

Technical Report # KU-EC-09-2: Extension of One-Dimensional Proximity Regions to Higher Dimensions

Elvan Ceyhan *

February 8, 2009

Abstract

Proximity maps and regions are defined based on the relative allocation of points from two or more classes in an area of interest and are used to construct random graphs called proximity catch digraphs (PCDs) which have applications in various fields. The simplest of such maps is the spherical proximity map which maps a point from the class of interest to a disk centered at the same point with radius being the distance to the closest point from the other class in the region. The spherical proximity map gave rise to class cover catch digraph (CCCD) which was applied to pattern classification. Furthermore for uniform data on the real line, the exact and asymptotic distribution of the domination number of CCCDs were analytically available. In this article, we determine some appealing properties of the spherical proximity map in compact intervals on the real line and use these properties as a guideline for defining new proximity maps in higher dimensions. Delaunay triangulation is used to partition the region of interest in higher dimensions. Furthermore, we introduce the auxiliary tools used for the construction of the new proximity maps, as well as some related concepts that will be used in the investigation and comparison of them and the resulting graphs. We characterize the geometry invariance of PCDs for uniform data. We also provide some newly defined proximity maps in higher dimensions as illustrative examples.

Keywords: class cover catch digraph (CCCD); Delaunay triangulation; domination number; proximity map; proximity catch digraph; random graph; relative density; triangle center

*Address: Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey. e-mail: elceyhan@ku.edu.tr, tel:+90 (212) 338-1845, fax: +90 (212) 338-1559.

1 Introduction

Classification and clustering have received considerable attention in the statistical literature. In recent years, a new classification approach has been developed. This approach is based on the proximity maps that incorporate the relative positions of the data points from various classes. Proximity maps and the associated (di)graphs are used in disciplines where shape and structure are crucial. Examples include computer vision (dot patterns), image analysis, pattern recognition (prototype selection), geography and cartography, visual perception, biology, etc. *Proximity graphs* were first introduced by Toussaint (1980), who called them *relative neighborhood graphs*. The notion of relative neighborhood graph has been generalized in several directions and all of these graphs are now called proximity graphs. From a mathematical and algorithmic point of view, proximity graphs fall under the category of *computational geometry*.

A general definition of proximity graphs is as follows: Let V be any finite or infinite set of points in \mathbb{R}^d . Each (unordered) pair of points $(p, q) \in V \times V$ is associated with a neighborhood $\mathfrak{N}(p, q) \subseteq \mathbb{R}^d$. Let \mathfrak{P} be a property defined on $\mathfrak{N} = \{\mathfrak{N}(p, q) : (p, q) \in V \times V\}$. A *proximity* (or *neighborhood*) *graph* $G_{\mathfrak{N}, \mathfrak{P}}(V, E)$ defined by the property \mathfrak{P} is a graph with the vertex set V and the edge set E such that $(p, q) \in E$ iff $\mathfrak{N}(p, q)$ satisfies property \mathfrak{P} . Examples of most commonly used proximity graphs are the Delaunay tessellation, the boundary of the convex hull, the Gabriel graph, relative neighborhood graph, Euclidean minimum spanning tree, and sphere of influence graph of a finite data set. See, e.g., Jaromczyk and Toussaint (1992) for more detail. The *Delaunay tessellation* $\mathcal{D}(V)$ of a finite set of points V , is the dual of the Voronoi diagram generated by V . See Okabe et al. (2000) for further details. The *convex hull* of a set $V \subset \mathbb{R}^d$, denoted as $\mathcal{C}_H(V)$, is the intersection of all convex sets (there exists infinitely many of them) that contain V . The boundary of $\mathcal{C}_H(V)$ can be viewed as a proximity graph which is also a subgraph of $\mathcal{D}(V)$. The *Gabriel graph* of V , denoted as $GG(V)$, is defined as the graph in which (p, q) is an edge of $GG(V)$ iff the circle centered at the midpoint of the line segment \overline{pq} and with diameter $d(p, q)$ (the distance between p and q) does not contain any other points from V . The *relative neighborhood graph* of V is a prominent representative of the family of graphs which are defined by some sort of neighborliness. For a set of points $V \subset \mathbb{R}^d$, the relative neighborhood graph of V , denoted $RNG(V)$, is a graph with vertex set V and edge set which are exactly the pairs (p, q) of points for which $d(p, q) \leq \min_{v \in V} \max(d(p, v), d(q, v))$. That is, (p, q) is an edge of $RNG(V)$ iff $\text{Lune}(p, q)$ does not contain any other points of V , where $\text{Lune}(p, q)$ is defined as the intersection of two discs centered at p, q each with radius $d(p, q)$ (see, e.g., Jaromczyk and Toussaint (1992)). The *Euclidean minimum spanning tree* of V , denoted $EMST(V)$, is defined as the spanning tree in which the sum of the Euclidean lengths of the edges yield the minimum over all spanning trees with vertex set V . The *sphere of influence graph* on V , denoted as $SIG(V)$, has vertex set V and (p, q) as an edge iff the circles centered at p and q with radii $\min_{v \in V \setminus \{p\}} d(p, v)$ and $\min_{v \in V \setminus \{q\}} d(q, v)$, respectively, have nonempty intersection. Note that $EMST(V)$ is a subgraph of $RNG(V)$ which is a subgraph of $GG(V)$ which is a subgraph of $\mathcal{D}(V)$ (see Okabe et al. (2000)). Furthermore, in the examples above, $d(x, y)$, can be any distance in \mathbb{R}^d . Furthermore, the distance between a point x and a set A is defined as $d(x, A) := \inf_{y \in A} d(x, y)$; and the distance between two sets A and B is defined as $d(A, B) := \inf_{(x, y) \in A \times B} d(x, y)$.

A *digraph* is a directed graph, i.e., a graph with directed edges from one vertex to another based on a binary relation. Then the pair $(p, q) \in V \times V$ is an ordered pair and is an *arc* (directed edge) denoted as pq to reflect the difference between an arc and an edge. For example, the nearest neighbor (di)graph in Paterson and Yao (1992) is a proximity digraph. The *nearest neighbor digraph*, denoted as $NND(V)$, has the vertex set V and pq as an arc iff $d(p, q) = \min_{v \in V \setminus \{p\}} d(p, v)$. That is, pq is an arc of $NND(V)$ iff q is a nearest neighbor of p . Note that if pq is an arc in $NND(V)$, then (p, q) is an edge in $RNG(V)$.

The *proximity catch digraphs* (PCDs) are based on the property \mathfrak{P} that is determined by the following mapping which is defined in a more general space than \mathbb{R}^d . Let (Ω, \mathcal{M}) be a measurable space. The *proximity map* $N(\cdot)$ is given by $N : \Omega \rightarrow \wp(\Omega)$, where $\wp(\cdot)$ is the power set functional, and the *proximity region* of $x \in \Omega$, denoted as $N(x)$, is the image of $x \in \Omega$ under $N(\cdot)$. The points in $N(x)$ are thought of as being “closer” to $x \in \Omega$ than are the points in $\Omega \setminus N(x)$. Proximity maps are the building blocks of the *proximity graphs* of Toussaint (1980); an extensive survey is available by Jaromczyk and Toussaint (1992). The PCD

D has the vertex set $\mathcal{V} = \{p_1, \dots, p_n\}$ and the arc set \mathcal{A} is defined as $p_i p_j \in \mathcal{A}$ iff $p_j \in N(p_i)$ for $i \neq j$. Notice that the PCD D depends on the *proximity* map $N(\cdot)$, and if $p_j \in N(p_i)$, then $N(p_i)$ is said to *catch* p_j . Hence the name *proximity catch digraph*. If arcs of the form $p_i p_i$ (i.e., loops) were allowed, D would have been called a *pseudodigraph* according to some authors (see, e.g., Chartrand and Lesniak (1996)).

The finite and asymptotic distribution of the domination number of CCCDs for uniform data in \mathbb{R} are mathematically tractable. In this article, we determine some appealing properties of the proximity map associated with CCCD for data in a compact interval in \mathbb{R} and use them as guidelines for defining new proximity maps in higher dimensions. As CCCD behaves nicely for uniform data in \mathbb{R} (in the sense that the exact and asymptotic distributions of the domination number and relative arc density are available), by emulating its properties in higher dimensions we expect the new PCDs will behave similarly. Furthermore, we introduce some auxiliary tools used for the construction of the new proximity maps, as well as some related concepts that will be used in the investigation and comparison of the proximity maps. Additionally, we discuss the conditions for the geometry invariance for uniform data in triangles.

We describe the data-random PCDs in Section 2, Voronoi diagrams and Delaunay tessellations in Section 3, the appealing properties of spherical proximity maps in \mathbb{R} in Section 4, transformations preserving uniformity on triangles in \mathbb{R}^2 in Section 5, triangle centers in Section 6, vertex and edge regions in Section 7, and proximity regions in Delaunay tessellations in Section 8. We present the results on relative arc density and the domination number of the new PCDs in Section 9, introduce two new proximity maps in Section 10, and discussion and conclusions in Section 11.

2 Data-Random Proximity Catch Digraphs

Priebe et al. (2001) introduced the class cover catch digraphs (CCCDs) and gave the exact and the asymptotic distribution of the domination number of the CCCD based on two classes \mathcal{X}_n and \mathcal{Y}_m both of which are random samples from uniform distribution on a compact interval in \mathbb{R} . DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a), Priebe et al. (2003b), and DeVinney and Priebe (2006) applied the concept in higher dimensions and demonstrated relatively good performance of CCCDs in classification. The methods employed involve *data reduction (condensing)* by using approximate minimum dominating sets as *prototype sets* since finding the exact minimum dominating set is in general an NP-hard problem — in particular, for CCCDs — (see DeVinney (2003)). Furthermore, the exact and the asymptotic distribution of the domination number of the CCCDs are not analytically tractable in dimensions greater than 1.

Let $\mathcal{X}_n = \{X_1, \dots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, \dots, Y_m\}$ be two sets of Ω -valued random variables from classes \mathcal{X} and \mathcal{Y} whose joint pdf is $f_{X,Y}$. Let $d(\cdot, \cdot) : \Omega \times \Omega \rightarrow [0, \infty)$ be a distance function. The class cover problem for a target class, say \mathcal{X}_n , refers to finding a collection of neighborhoods, $N(X_i)$ around $X_i \in \mathcal{X}_n$ such that (i) $\mathcal{X}_n \subseteq (\bigcup_i N(X_i))$ and (ii) $\mathcal{Y}_m \cap (\bigcup_i N(X_i)) = \emptyset$. A collection of neighborhoods satisfying both conditions is called a *class cover*. A cover satisfying condition (i) is a *proper cover* of class \mathcal{X} while a collection satisfying condition (ii) is a *pure cover* relative to class \mathcal{Y} . From a practical point of view, for example for classification, of particular interest are the class covers satisfying both (i) and (ii) with the smallest collection of neighborhoods, i.e., minimum cardinality cover. This class cover problem is a generalization of the set cover problem in Garfinkel and Nemhauser (1972) that emerged in statistical pattern recognition and machine learning, where an edited or condensed set (i.e., prototype set) is selected from \mathcal{X}_n (see, e.g., Devroye et al. (1996)).

In particular, the proximity regions are constructed using data sets from two classes. Given $\mathcal{Y}_m \subseteq \Omega$, the *proximity map* $N_{\mathcal{Y}}(\cdot) : \Omega \rightarrow \wp(\Omega)$ associates a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$ with each point $x \in \Omega$. The region $N_{\mathcal{Y}}(x)$ is defined in terms of the distance between x and \mathcal{Y}_m . More specifically, our proximity maps will be based on the relative position of points from class \mathcal{X} with respect to the Delaunay tessellation of the points from class \mathcal{Y} . See Okabe et al. (2000) for more on Delaunay tessellations.

If $\mathcal{X}_n = \{X_1, \dots, X_n\}$ is a set of Ω -valued random variables then $N_{\mathcal{Y}}(X_i)$ are random sets. If X_i are independent identically distributed then so are the random sets $N_{\mathcal{Y}}(X_i)$. The data-random PCD D — associated with $N_{\mathcal{Y}}(\cdot)$ — is defined with vertex set $\mathcal{X}_n = \{X_1, \dots, X_n\}$ and arc set \mathcal{A} by $X_i X_j \in \mathcal{A} \iff X_j \in N_{\mathcal{Y}}(X_i)$. Since this relationship is not symmetric, a digraph is used rather than a graph. The random digraph D depends on the (joint) distribution of the X_i and on the map $N_{\mathcal{Y}}(\cdot)$. Let $\mu(N_{\mathcal{Y}}) := P(X_i X_j \in \mathcal{A}) = P(X_j \in N_{\mathcal{Y}}(X_i))$; so $\mu(N_{\mathcal{Y}})$ is the probability of having an arc from X_i to X_j , hence is called *arc probability* for the PCD based on $N_{\mathcal{Y}}$.

The PCDs are closely related to the *proximity graphs* of Jaromczyk and Toussaint (1992) and might be considered as a special case of *covering sets* of Tuza (1994) and *intersection digraphs* of Sen et al. (1989). This data random proximity digraph is a *vertex-random proximity digraph* which is not of standard type (see, e.g., Janson et al. (2000)). The randomness of the PCDs lies in the fact that the vertices are random with joint pdf $f_{X,Y}$, but arcs $X_i X_j$ are deterministic functions of the random variable X_j and the set $N_{\mathcal{Y}}(X_i)$. For example, the CCCD of Priebe et al. (2001) can be viewed as an example of PCD with $N_{\mathcal{Y}}(x) = B(x, r(x))$, where $r(x) := \min_{y \in \mathcal{Y}_m} d(x, y)$. The CCCD is the digraph of order n with vertex set \mathcal{X}_n and an arc from X_i to X_j iff $X_j \in B(X_i, r(X_i))$. That is, there is an arc from X_i to X_j iff there exists an open ball centered at X_i which is “pure” (or contains no elements) of \mathcal{Y}_m , and simultaneously contains (or “catches”) point X_j .

Notice that the CCCDs are defined with (open) balls only, whereas PCDs are not based on a particular geometric shape or a functional form; that is, PCDs admit $N_{\mathcal{Y}}(\cdot)$ to be any type of region, e.g., circle (ball), arc slice, triangle, a convex or nonconvex polygon, etc. In this sense, the PCDs are defined in a more general setting compared to CCCDs. On the other hand, the types of PCDs introduced in this article are well-defined for points restricted to the convex hull of \mathcal{Y}_m , $\mathcal{C}_H(\mathcal{Y}_m)$. Moreover, the proximity maps introduced in this article will yield closed regions. Furthermore, the CCCDs based on balls use proximity regions that are defined by the obvious metric, while the PCDs do not necessarily require a metric, but some sort of dissimilarity measure only.

3 Voronoi Diagrams and Delaunay Tessellations

The proximity map in \mathbb{R} defined as $B(X, r(X))$ where $r(X) = \min_{Y \in \mathcal{Y}_m} d(X, Y)$ in (Priebe et al. (2001)). Our next goal is to extend this idea to higher dimensions and investigate the properties of the associated digraph. Now let $\mathcal{Y}_m = \{y_1, \dots, y_m\} \subset \mathbb{R}$ and $Y_{j:m}$ be the j^{th} order statistic. The above definition of the proximity map is based on the intervals $I_j = (Y_{(j-1):m}, Y_{j:m})$ for $j = 0, \dots, (m+1)$ with $y_{0:m} = -\infty$ and $y_{(m+1):m} = \infty$. This *intervalization* can be viewed as a tessellation since it partitions $\mathcal{C}_H(\mathcal{Y}_m)$, the convex hull of \mathcal{Y}_m . For $d > 1$, a natural tessellation that partitions $\mathcal{C}_H(\mathcal{Y}_m)$ is the Delaunay tessellation, where each Delaunay cell is a $(d+1)$ -simplex (e.g., a Delaunay cell is an interval for $d = 1$, a triangle for $d = 2$, a tetrahedron for $d = 3$, and so on.) Let \mathcal{T}_j be the j^{th} Delaunay cell for $j = 1, \dots, J$ in the Delaunay tessellation of \mathcal{Y}_m . In \mathbb{R} , the cell that contains x is implicitly used to define the proximity map.

A *tessellation* is a partition of a space into convex polytopes; tessellation of the plane (into convex polygons) is the most frequently studied case (Schoenberg (2002)). Given $2 \leq n < \infty$ distinct points in \mathbb{R}^2 , we associate all points in the space with the closest member(s) of the point set with respect to the Euclidean distance. The result is a tessellation of the plane into a set of regions associated with the n points. We call this tessellation the *planar ordinary Voronoi diagram* generated by the point set and the regions *ordinary Voronoi polygons*. See Figure 1 for an example with $m = |\mathcal{Y}_m| = 10$ \mathcal{Y} points iid from $\mathcal{U}((0, 1) \times (0, 1))$.

In general, let $P = \{p_1, p_2, \dots, p_n\}$ be n points in \mathbb{R}^d where $2 \leq n < \infty$ and $p_i \neq p_j$ for $i \neq j$, $i, j \in [n] := \{1, 2, \dots, n\}$ and let $\|\cdot\|$ denote the norm functional. We call the region $\mathcal{V}_C(p_i) = \{x \in \mathbb{R}^d : \|x - p_i\| \leq \|x - p_j\| \text{ for } j \neq i, j \in [n]\}$ the (ordinary) *Voronoi polygon* or *cell* associated with p_i and the set $\mathfrak{V} = \{\mathcal{V}_C(p_1), \dots, \mathcal{V}_C(p_n)\}$ the *Voronoi diagram* or *Dirichlet tessellation* generated by P . We call p_i the *generator* of $\mathcal{V}_C(p_i)$. The Voronoi diagram partitions the space into disjoint regions (which are also called *tiles* or *Thiessen polygons* in \mathbb{R}^2). Notice that we still say \mathfrak{V} *partitions* the space \mathbb{R}^d , although $\mathcal{V}_C(p_j)$

are not necessarily disjoint, but if nonempty the intersection lies in a lower dimension, or equivalently, has zero \mathbb{R}^d -Lebesgue measure. We stick to this convention throughout the article. The intersection of two Voronoi cells, if nonempty, i.e., for $i \neq j$, $\mathcal{V}_C(p_i) \cap \mathcal{V}_C(p_j) \neq \emptyset$, is called a *Voronoi edge*. If a Voronoi edge is not a point, then $\mathcal{V}_C(p_i)$ and $\mathcal{V}_C(p_j)$ are said to be *adjacent*. An end point of a Voronoi edge is called a *Voronoi vertex*. \mathfrak{V} is called *degenerate* if at a Voronoi vertex, more than three Voronoi polygons intersect, and *non-degenerate* otherwise. A detailed discussion including the history of Voronoi diagrams is available in Okabe et al. (2000).

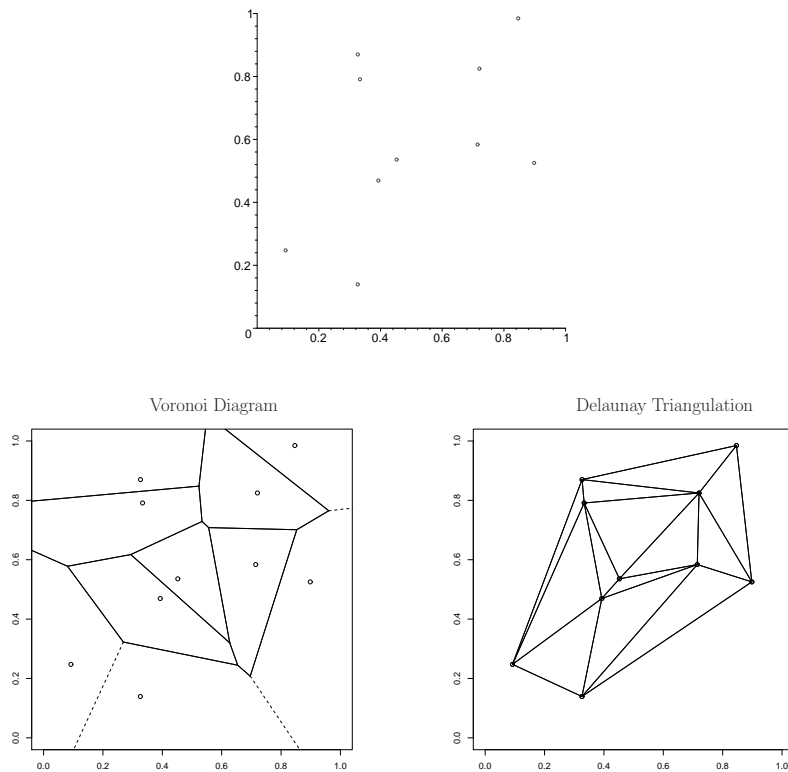


Figure 1: Depicted are 10 \mathcal{Y} points generated iid $\mathcal{U}(0, 1) \times (0, 1)$ (top), the corresponding Voronoi diagram (bottom left) and the Delaunay triangulation (bottom right).

Given a Voronoi diagram with $n \geq d + 1$ non-coplanar (i.e., not all the points lie on a $(d - 1)$ -dimensional hyperplane) generators, P , in \mathbb{R}^d , we join all pairs of generators whose Voronoi cells have a common Voronoi edge. The resulting tessellation is called the *Delaunay tessellation* of P . See Figure 1 for the Delaunay triangulation associated with the Voronoi diagram based on 10 points iid from $\mathcal{U}((0, 1) \times (0, 1))$. By definition a Delaunay tessellation of a finite set of points, P , is the dual of the *Voronoi diagram* based on the same set. The tessellation yields a (unique) polytopization provided that no more than $(d + 1)$ points in \mathbb{R}^d are cospherical (i.e., no more than $(d + 1)$ points lie on the boundary of a (hyper)sphere in \mathbb{R}^d). Moreover, the circumsphere of each Delaunay polytope (i.e., the sphere that contains the vertices of the Delaunay polytope on its boundary) is pure from the set P ; i.e., the interior of the circumsphere of the Delaunay polytope does not contain any points from P . The Delaunay tessellation partitions $\mathcal{C}_H(P)$. In particular, in \mathbb{R}^2 , the tessellation is a *triangulation* that yields triangles T_j , $j = 1, \dots, J$ (see, e.g., Okabe et al. (2000)) provided that no more than three points are cocircular (i.e., no more than three points lie on the boundary of some circle in \mathbb{R}^2). In this article we adopt the convention that a triangle refers to the closed region bounded by its edges. See Figure 2 for two examples: an example with $n = 200$ \mathcal{X} points $\stackrel{iid}{\sim} \mathcal{U}((0, 1) \times (0, 1))$, the

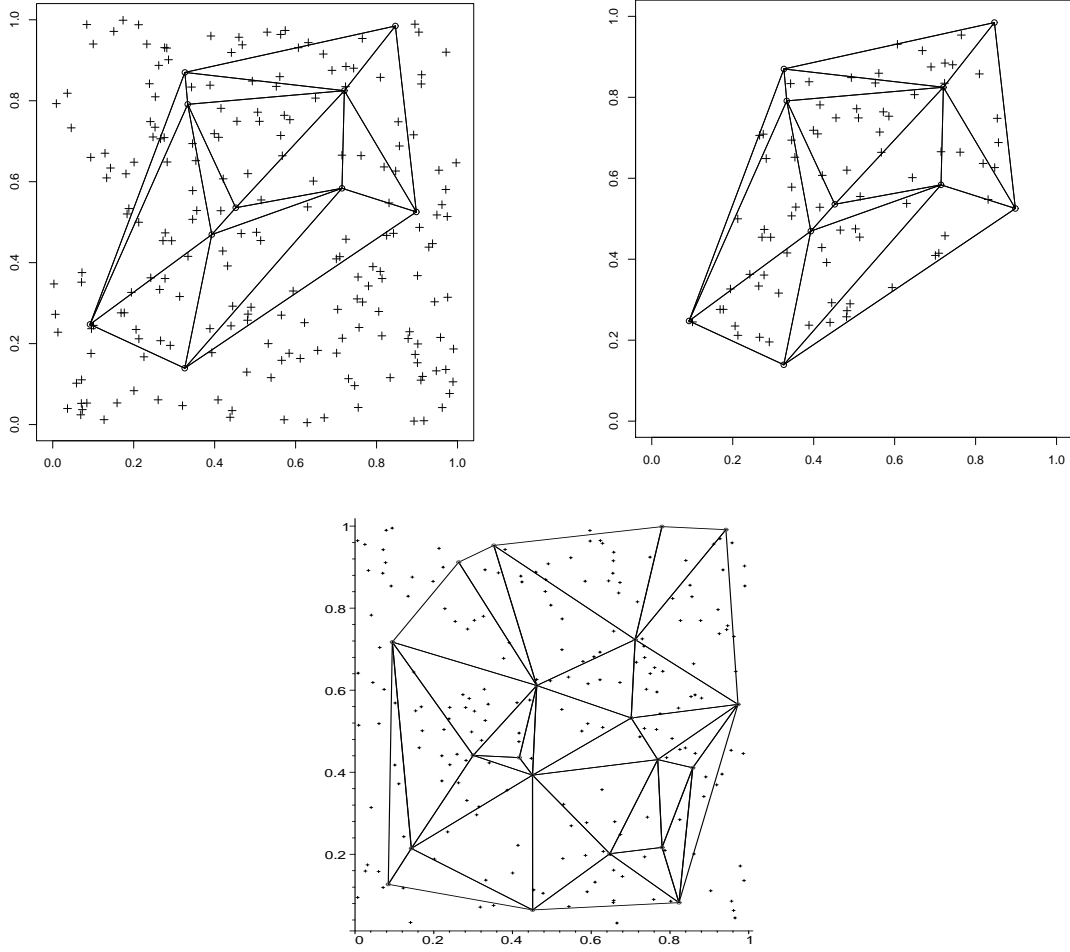


Figure 2: A realization of 200 X points and the Delaunay triangulation based on 10 \mathcal{Y} points in Figure 1 (top). Another realization of 200 X and 20 \mathcal{Y} points and the Delaunay triangulation based on \mathcal{Y}_m (bottom).

uniform distribution on the unit square and the Delaunay triangulation based on the 10 \mathcal{Y} points in Figure 1; and an example with $n = 200$ X points $\stackrel{iid}{\sim} \mathcal{U}((0, 1) \times (0, 1))$, the uniform distribution on the unit square and the Delaunay triangulation is based on 20 \mathcal{Y} points also $\stackrel{iid}{\sim} \mathcal{U}((0, 1) \times (0, 1))$.

3.1 Poisson Delaunay Triangles

The Delaunay triangles are based on a given set of points \mathcal{Y}_m . The set \mathcal{Y}_m can be assumed to come from a Poisson point process on a finite region, and in the application of PCDs, to remove the conditioning on \mathcal{Y}_m , it is suggested that \mathcal{Y}_m comes from a Poisson point process for prospective research directions. We briefly describe the Poisson point processes and Poisson Delaunay triangles.

A *stochastic point process* on \mathbb{R}^d is defined to be a process in which points are generated according to a probability distribution $P(|\mathcal{Y} \cap B| = k)$, $k = 0, 1, 2, \dots$, over any $B \subseteq \mathbb{R}^d$. For example, a *binomial point process* is a stochastic point process in which n points are generated over a bounded set $S \subsetneq \mathbb{R}^d$ according to

the uniform distribution. In particular, if $d = 2$ the process is called a *planar stochastic point process*. If two points coincide with probability zero, then it is a *simple stochastic point process*. A stochastic point process is said to be *locally finite* if any finite region $B \subset \mathbb{R}^d$ contains a finite number of points with probability 1 under the process (see Okabe et al. (2000)).

We have built the Delaunay tessellation using \mathcal{Y}_m with finite sample size. Suppose \mathcal{Y} is from a stochastic point process. One of the most fundamental locally finite stochastic point processes is the *Poisson point process*, which is defined as the process that satisfies

$$P(|\mathcal{Y} \cap B| = k) = \frac{\lambda \cdot V(B) \cdot e^{-\lambda \cdot V(B)} \cdot k}{k!}, \quad k = 0, 1, 2, \dots$$

for any $B \subseteq \mathbb{R}^d$, where $V(\cdot)$ denotes the d -dimensional volume functional and $\lambda > 0$ is the intensity (number of points per unit volume) of the process. We can also define the Poisson point process as the limit of the binomial point process in the sense of expanding the finite region S to an infinite region while keeping $\lambda = n/V(S)$ constant. We call the Delaunay tessellation based on a finite data set from a Poisson point process *Poisson Delaunay tessellation* and denote it \mathcal{D}_P . The associated Voronoi diagram is called the *Poisson Voronoi diagram* and is denoted by \mathcal{V}_P . For more detail on the properties of \mathcal{V}_P , see, Okabe et al. (2000).

A *simplex* in \mathbb{R}^d is the convex hull of any $(d+1)$ points in general position, i.e., no $(d+1)$ of the points lie in a $(d-1)$ -dimensional hyperplane in \mathbb{R}^d . The simplex is the point itself for $d = 0$, the line segment joining the two points for $d = 1$, a triangle for $d = 2$, a tetrahedron for $d = 3$, and so on. Each Poisson Delaunay cell of \mathcal{D}_P is a $(d+1)$ -dimensional simplex whose vertices are x_0, x_1, \dots, x_m from the Poisson point process. Any s -face of \mathcal{D}_P is an $(s+1)$ -dimensional simplex with vertices x_0, \dots, x_s , also points from the Poisson point process. There are $\binom{d+s}{s+1}$ many s -faces contained in a Poisson Delaunay cell for $0 \leq s \leq d$.

Let c and r be the circumcenter and circumradius, respectively, of a $(d+1)$ -dimensional Poisson Delaunay cell in \mathbb{R}^d . Then the $(d+1)$ vertices of the cell are the points $\{c + r u_i\}$ where $\{u_i\}$ are the unit vectors for $i = 0, 1, \dots, d$. The ergodic joint probability density function (pdf) of \mathcal{D}_P is completely specified as

$$f(r, u_0, \dots, u_d) = a(\lambda, d) \Delta_d r^{d^2-1} \exp(-\lambda w_d r^d)$$

where Δ_d is the volume of the $(d+1)$ -simplex with vertices u_0, \dots, u_d , $w_d = \pi^{d/2}/\Gamma(d/2 + 1)$ and

$$a(\lambda, d) = \frac{\pi^{(d^2+1)/2} \Gamma(m^2/2) [2\lambda \Gamma((d+1)/2)]^d}{d^{d-2} \Gamma(d/2)^{2d+1} \Gamma((d^2+1)/2)}.$$

The circumradius r may be viewed as a measure of size of the $(d+1)$ -simplex, and is independent of $\{u_0, \dots, u_d\}$. The pdf of r is a generalized gamma function with $t = d$, $q = d^2$ and $b = \lambda_d$, where a 3-parameter generalized gamma function is

$$f(x) = r b^{q/t} x^{q-1} \exp(-b x^t) / \Gamma(q(t)) \quad \text{for } x, t, p, q > 0.$$

Let V_d denote the volume of a typical Poisson Delaunay cell. Then for $d = 2$, the expected value of the k^{th} moment of volume of a typical Poisson Delaunay triangle is

$$\mathbf{E}[V_2^k] = \frac{\Gamma((3k+5)/2) \Gamma(k/2+1)}{3 \Gamma((k+3)/2)^2 2^k \pi^{k-1/2} \lambda^k}, \quad k = 1, 2, \dots$$

In \mathbb{R}^2 , the joint pdf of a pair of inner angles arbitrarily selected from an arbitrary triangle in \mathcal{D}_P is given by

$$f(x, y) = \frac{8}{3\pi} (\sin x)(\sin y) \sin(x+y), \quad \text{for } x, y > 0 \text{ and } x+y < \pi.$$

Notice that the mode of this density is at $x = y = \pi/3$, which implies that the most frequent triangles in a \mathcal{D}_P are nearly equilateral triangles. By integrating over y , we obtain the pdf of a randomly selected inner angle of an arbitrary triangle from \mathcal{D}_P :

$$f(x) = \left[\frac{4}{3\pi} ((\pi - x) \cos x + \sin x) \sin x \right] \mathbf{I}(0 < x < \pi)$$

where $\mathbf{I}(\cdot)$ is the indicator function. Then the expected value of X is $\mathbf{E}[X] = \pi/3$ and $\mathbf{E}[X^2] = 2\pi^2/9 - 5/6$.

Mardia et al. (1977) derived the pdf of the minimum angle as

$$f_1(x) = \left(\frac{2}{\pi} (\pi - 3x) \sin 2x + \cos 2x - \cos 4x \right) \mathbf{I}\left(0 < x < \frac{\pi}{3}\right),$$

and Boots (1986) gave the pdf of maximum angle,

$$f_3(x) = \left[\frac{2}{\pi} (3x(\sin 2x) - \cos 2x + \cos 4x - \pi \sin 2x) \right] \mathbf{I}\left(\frac{\pi}{3} < x < \frac{\pi}{2}\right) + \left[\frac{1}{\pi} (4\pi(\cos x)(\sin x) + 3 \sin x^2 - \cos x^2 - 4x(\cos x)(\sin x) + 1) \right] \mathbf{I}\left(\frac{\pi}{2} < x < \pi\right).$$

The distribution of the length of an arbitrary edge e of an arbitrary triangle from \mathcal{D}_P is

$$f_L(x) = \left(\pi \lambda \frac{x}{3} \right) \left(\sqrt{\lambda} x \exp\left(-\pi \lambda \frac{x^2}{4}\right) + \operatorname{erfc}\left(\sqrt{\pi \lambda} \frac{x}{2}\right) \right)$$

where erfc is the complimentary error function $\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt = \pi^{-\frac{1}{2}} \Gamma\left(\frac{1}{2}, z^2\right)$ (Okabe et al. (2000)).

4 The Appealing Properties of Spherical Proximity Maps in \mathbb{R}

The CCCDs have desirable properties such as having the finite sample and asymptotic distributions of the domination number available. In this section, we determine some appealing properties of the proximity map associated with CCCD for uniform data in a compact interval in \mathbb{R} and use them as guidelines for defining new proximity maps in higher dimensions. We believe these properties cause the CCCD to behave so “nicely” in \mathbb{R} and the more they are satisfied by the new PCDs in higher dimensions, the more likely the new PCDs to have similar behaviour. Furthermore, we introduce the auxiliary tools used for the construction of the new proximity maps, as well as some related concepts that will be used in the investigation and comparison of the proximity maps.

Let $\mathcal{Y}_m = \{y_1, \dots, y_m\} \subset \mathbb{R}$. Then the proximity map associated with CCCD is defined as the open ball $N_S(x) := B(x, r(x))$ for all $x \in \mathbb{R} \setminus \mathcal{Y}_m$, where $r(x) = \min_{y \in \mathcal{Y}_m} d(x, y)$ (see Section 2 and Priebe et al. (2001)) with $d(x, y)$ being the Euclidean distance between x and y . For $x \in \mathcal{Y}_m$, define $N_S(x) = \{x\}$. Notice that a ball is a sphere in higher dimensions, hence the name *spherical proximity map* and the notation N_S . Furthermore, dependence on \mathcal{Y}_m is through $r(x)$. Note that, this proximity map is based on the intervals $I_i = (y_{(i-1):m}, y_{i:m})$ for $i = 0, \dots, (m+1)$ with $y_{0:m} = -\infty$ and $y_{(m+1):m} = \infty$ where $y_{i:m}$ is the i^{th} order statistic of \mathcal{Y}_m .

For $X_i \stackrel{iid}{\sim} \mathcal{U}(I_i)$, without loss of generality we can assume $I_i = (0, 1)$. Then the arc probability $\mu(N_S) = P(X_2 \in N_S(X_1)) = 1/2$, since $P(X_2 \in N_S(X_1)) = \int_0^{1/2} \int_0^{2x_1} dx_2 dx_1 + \int_{1/2}^1 \int_{2-2x_1}^1 dx_2 dx_1 = 1/2$.

A natural extension of the proximity region $N_S(x)$ to multiple dimensions (i.e., to \mathbb{R}^d with $d > 1$) is obtained by the same definition as above; that is, $N_S(x) := B(x, r(x))$ where $r(x) := \min_{y \in \mathcal{Y}_m} d(x, y)$. The spherical proximity map $N_S(x)$ is well-defined for all $x \in \mathbb{R}^d$ provided that $\mathcal{Y}_m \neq \emptyset$. Extensions to \mathbb{R}^2 and higher dimensions with the spherical proximity map — with applications in classification — are investigated in DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a), Priebe et al. (2003b), and DeVinney and Priebe (2006). However, finding the minimum dominating set of the PCD associated with $N_S(\cdot)$ is an NP-hard problem (DeVinney (2003)) and the distribution of the domination number is not analytically tractable for $d > 1$ (Ceyhan (2004)). This drawback has motivated the definition of new types of proximity maps in higher dimensions. Note that for $d = 1$, such problems do not exist. We state some appealing properties of the proximity map $N_S(x) = B(x, r(x))$ in \mathbb{R} and use them as guidelines for our definition of new proximity maps:

- P1** $N_S(x)$ is well-defined for all $x \in \mathcal{C}_H(\mathcal{Y}_m) = [y_{1:m}, y_{m:m}]$.
- P2** $x \in N_S(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y}_m)$.
- P3** The point x is at the *center* of $N_S(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y}_m)$.
- P4** For $x \in I_i \subseteq \mathcal{C}_H(\mathcal{Y}_m)$, $N_S(x)$ and I_i are of the *same type*; i.e., they are both intervals.
- P5** For $x \in I_i \subseteq \mathcal{C}_H(\mathcal{Y}_m)$, $N_S(x)$ mimics the shape of I_i ; i.e., it is (geometrically) *similar* to I_i .
- P6** For $x \in I_i$, $N_S(x)$ is a proper subset of I_i for all $x \in I_i \setminus \{(y_{(i-1):m} + y_{i:m})/2\}$ (or almost everywhere in I_i).
- P7** For $x \in I_i$ and $y \in I_j$ with $i \neq j$, $N_S(x)$ and $N_S(y)$ are disjoint regions.
- P8** The size (i.e., measure) of $N_S(x)$ is continuous in x ; that is, for each $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that $||N_S(y)| - |N_S(x)|| < \varepsilon$ whenever $|d(x, y)| < \delta(\varepsilon)$.
- P9** The arc probability $\mu(N_S)$ does not depend on the support interval for uniform data in \mathbb{R} .

Notice that properties **P1**, **P2**, and **P3** also hold for all $x \in \mathbb{R}$. **P9** implies that not only the arc probability but also the distribution of the relative arc density and domination number do not depend on the support interval either. This independence of the support set is called *geometry invariance* in higher dimensions (see Section 5). For N_S it suffices to work with $\mathcal{U}(0, 1)$ data, and in higher dimensions we will be able to consider only uniform data in an equilateral triangle for PCDs based on proximity maps that satisfy **P9**.

Suppose we partition the convex hull of $\mathcal{Y}_m, \mathcal{C}_H(\mathcal{Y}_m)$ by Delaunay tessellation. Let \mathcal{T}_j be the j^{th} Delaunay cell in the Delaunay tessellation of \mathcal{Y}_m for $j = 1, \dots, J$, where J is the total number of Delaunay cells. See Figure 2 for two \mathcal{Y} sets of sizes 10 and 20 and the corresponding Delaunay triangulations.

Note that **P4** and **P5** are equivalent when $d = 1$ for $x \in \mathcal{C}_H(\mathcal{Y}_m)$, since any two (compact) intervals in \mathbb{R} are (geometrically) similar. For $d > 1$, **P5** implies **P4** only, since, for example, for $d = 2$, any two triangles are not necessarily similar, but similar triangles are always of the same type; they are triangles.

Notice that $N_S(\cdot)$ satisfies only **P1**, **P2**, **P3**, and **P8** in \mathbb{R}^d with $d > 1$. **P4** and **P5** fail since $N_S(x)$ is a sphere for $x \notin \mathcal{Y}_m$, but \mathcal{T}_j is a $(d + 1)$ -simplex. For any $x \in \mathcal{T}_j \subset \mathbb{R}^d$, $B(x, r(x)) \not\subset \mathcal{T}_j$, so **P6** also fails, furthermore this also implies that $N_S(x)$ and $N_S(y)$ might overlap for x, y from two distinct cells, hence **P7** is violated. The arc probability $\mu(N_S)$ depends on the support set \mathcal{T}_i for $d > 1$ so **P9** is violated.

The appealing properties mentioned above can be extended to more general measurable spaces. Let (Ω, \mathcal{M}) be a measurable space, and let Ω_j , $j \in \{1, 2, \dots, J\}$ partition Ω , and μ be the associated measure on Ω . Then the appealing properties are

- P1** $N(x)$ is well-defined for all $x \in \Omega$.

P2 $x \in N(x)$ for all $x \in \Omega$.

P3 x is at the *center* of $N(x)$ for all $x \in \Omega$.

P4 For $x \in \Omega_j$, $N(x)$ and Ω_j are of the *same type*; they have the same functional form.

P5 For $x \in \Omega_j \subseteq \Omega$, $N(x)$ mimics the shape of Ω_j ; i.e., it is *similar* to Ω_j .

P6 For $X \in \Omega_j$, $N(X)$ is a proper subset of Ω_j a.s.

P7 For $x \in \Omega_j$ and $y \in \Omega_k$ with $j \neq k$, $N(x)$ and $N(y)$ are disjoint.

P8 The measure of $N(x)$ is continuous in x ; that is, for each $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that $|\mu(N(y)) - \mu(N(x))| < \delta(\varepsilon)$ whenever $\|y - x\| < \varepsilon$.

P9 The arc probability $\mu(N_S)$ does not depend on the support set for uniform data in Ω .

Property **P6** suggests a new concept.

Definition 4.1. The *superset region* for any proximity map $N(\cdot)$ in Ω is defined to be $\mathcal{R}_S(N) := \{x \in \Omega : N(x) = \Omega\}$. \square

For example, for $\Omega = I_i \subsetneq \mathbb{R}$, $\mathcal{R}_S(N_S) := \{x \in I_i : N_S(x) = I_i\} = \{(y_{(i-1):m} + y_{i:m})/2\}$, and for $\Omega = \mathcal{T}_i \subsetneq \mathbb{R}^d$, $\mathcal{R}_S(N_S) := \{x \in \mathcal{T}_i : N_S(x) = \mathcal{T}_i\}$. Note that for $x \in I_i$, $\lambda(N_S(x)) \leq \lambda(I_i)$ and $\lambda(N_S(x)) = \lambda(I_i)$ iff $x \in \mathcal{R}_S(N_S)$ where $\lambda(\cdot)$ is the Lebesgue measure on \mathbb{R} (also called \mathbb{R} -Lebesgue measure). So the proximity region of a point in $\mathcal{R}_S(N_S)$ has the largest \mathbb{R} -Lebesgue measure. Note that for $\mathcal{Y}_m = \{y_1, \dots, y_m\} \subset \mathbb{R}$, $\mathcal{R}_S(N_S) = \left\{ \frac{y_{1:m} + y_{2:m}}{2}, \dots, \frac{y_{(m-1):m} + y_{m:m}}{2} \right\}$. Note also that given \mathcal{Y}_m , $\mathcal{R}_S(N_S)$ is not a random set, but $\mathbf{I}(X \in \mathcal{R}_S(N_S))$ is a random variable. Furthermore, **P6** is equivalent to $\mathcal{R}_S(N_S)$ having zero \mathbb{R} -Lebesgue measure. On the other hand, for $x \in \partial(I_i) = \{y_{(i-1):m}, y_i\}$, the proximity region $N_S(x) = \{x\}$ has zero \mathbb{R} -Lebesgue measure. This suggests the following concept.

Definition 4.2. Let (Ω, μ) be a measurable space. The Λ_0 -region for any proximity map $N(\cdot)$ is defined to be $\Lambda_0(N) := \{x \in \Omega : \mu(N(x)) = 0\}$. \square

For $\Omega = \mathbb{R}^d$, $\Lambda_0(N_S) := \{x \in \mathbb{R}^d : \lambda(N_S(x)) = 0\}$. For example, for $\Omega = \mathcal{C}_H(\mathcal{Y}_m) \subsetneq \mathbb{R}$, $\Lambda_0(N_S) = \mathcal{Y}_m$, since $\lambda(N_S(x)) = 0$ iff $x \in \mathcal{Y}_m$.

Furthermore, given a set B of size n in $[y_{1:m}, y_{m:m}] \setminus \mathcal{Y}_m$, **P7** implies that the number of disconnected components in the PCD based on $N_S(\cdot)$ is at least the cardinality of $\{i \in [m] : B \cap I_i \neq \emptyset\}$, which is the set of indices of the intervals that contain some point(s) from B and $[m] := \{0, 1, \dots, m-1\}$.

5 Transformations Preserving Uniformity on Triangles in \mathbb{R}^2

The property **P9** suggests that in higher dimensions the arc probability of PCDs based on uniform data is geometry invariant, i.e., does not depend on the geometry of the support set. The set \mathcal{X}_n is assumed to be a set of iid uniform random variables on the convex hull of \mathcal{Y}_m ; i.e., a random sample from $\mathcal{U}(\mathcal{C}_H(\mathcal{Y}_m))$. In particular, conditional on $|\mathcal{X}_n \cap T_j| > 0$ being fixed, $\mathcal{X}_n \cap T_j$ will also be a set of iid uniform random variables on T_j for $j \in \{1, 2, \dots, J\}$. The geometry invariance property will reduce the triangle T_j as much as possible while preserving uniformity and the probabilities related to PCDs will simplify in notation and calculations. Below, we present such a transformation that reduces a single triangle to the standard equilateral triangle $T_e = T((0, 0), (1, 0), (1/2, \sqrt{3}/2))$.

Let $\mathcal{Y}_3 = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$ be three non-collinear points and $T(\mathcal{Y}_3)$ be the triangle with vertices y_1, y_2, y_3 . Let $X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$, the uniform distribution on $T(\mathcal{Y}_3)$, for $i = 1, \dots, n$. The pdf of $\mathcal{U}(T(\mathcal{Y}_3))$ is

$$f(u) = \frac{1}{A(T(\mathcal{Y}_3))} \mathbf{I}(u \in T(\mathcal{Y}_3)),$$

where $A(\cdot)$ is the area functional.

The triangle $T(\mathcal{Y}_3)$ can be carried into the first quadrant by a composition of transformations in such a way that the largest edge has unit length and lies on the x -axis, and the x -coordinate of the vertex nonadjacent to largest edge is less than $1/2$. We call the resultant triangle the *basic triangle* and denote it as T_b where $T_b = ((0, 0), (1, 0), (c_1, c_2))$ with $0 < c_1 \leq 1/2$, and $c_2 > 0$ and $(1 - c_1)^2 + c_2^2 \leq 1$. We will describe such transformations below: Let e_i be the edge opposite to the vertex y_i for $i \in \{1, 2, 3\}$. Find the lengths of the edges; say e_3 is of maximum length. Then scale the triangle so that e_3 is of unit length. Next translate y_1 to $(0, 0)$, and rotate (if necessary) the triangle so that $y_2 = (1, 0)$. If the y -coordinate of y_3 is negative reflect the triangle around the x -axis, then if x -coordinate of y_3 is greater than $1/2$, reflect it around $x = 1/2$, then the associated basic triangle T_b is obtained. So the basic triangle T_b can be obtained by a transformation denoted by ϕ_b which is a composition of rigid motion transformations (namely translation, rotation, and reflection) and scaling. Hence if $T(\mathcal{Y}_3)$ is transformed into T_b , then $T(\mathcal{Y}_3)$ is similar to T_b and $\phi_b(T(\mathcal{Y}_3)) = T_b$. Thus the random variables $X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ transformed along with $T(\mathcal{Y}_3)$ in the described fashion by ϕ_b satisfy $\phi_b(X_i) \stackrel{iid}{\sim} \mathcal{U}(T_b)$. So, without loss of generality, we can assume $T(\mathcal{Y}_3)$ to be the basic triangle, The functional form of T_b is

$$T_b = \{(x, y) \in \mathbb{R}^2 : y \geq 0; y \leq (c_2 x)/c_1; y \leq c_2(1 - x)/(1 - c_1)\}.$$

If $c_1 = 1/2$ and $c_2 = \sqrt{3}/2$, then T_b is an equilateral triangle; if $c_2 < \sqrt{c_1 - c_1^2}$, then T_b is an obtuse triangle; if $c_2 = \sqrt{c_1 - c_1^2}$, then T_b is a right triangle; and if $c_2 > \sqrt{c_1 - c_1^2}$, then T_b is an acute triangle. If $c_2 = 0$, then the T_b reduces to the unit interval $(0, 1)$. See Figure 3 for the domain of (c_1, c_2) for T_b and the part on which T_b is a non-acute triangle.

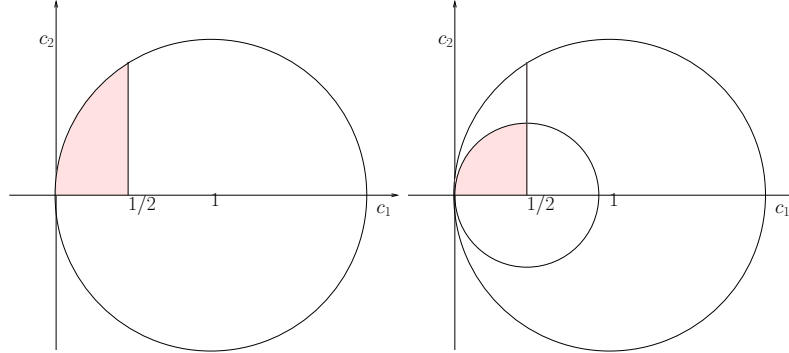


Figure 3: The shaded regions are the domain of (c_1, c_2) values for the basic triangle T_b (left) the domain where T_b is a non-acute triangle (right).

Lemma 5.1. *The arc probability $\mu(N_{\mathcal{Y}})$ of the PCD based on $N_{\mathcal{Y}}$ for uniform data on $T(\mathcal{Y}_3)$ is rigid-motion and scale invariant; i.e., $\mu(N_{\mathcal{Y}})$ does not change under rigid motion transformations and does not depend on the scale of the support triangle $T(\mathcal{Y}_3)$.*

Proof: We have shown that for $X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$, we have $\phi_b(X_i) \stackrel{iid}{\sim} \mathcal{U}(T_b)$ since $T(\mathcal{Y}_3)$ is similar to T_b . For uniform data, the set probabilities are calculated as the ratio of the area of the set to the total area. So

$P(X \in S \subseteq T(\mathcal{Y}_3)) = A(S)/A(T(\mathcal{Y}_3))$ and $P(\phi_b(X) \in \phi_b(S) \subseteq \phi_b(T(\mathcal{Y}_3))) = P(\phi_b(X) \in \phi_b(S) \subseteq T_b) = A(\phi_b(S))/A(T_b) = [kA(S)]/[kA(T(\mathcal{Y}_3))] = A(S)/A(T(\mathcal{Y}_3))$ where k is the scaling factor. Letting $X = X_j$ and $S = N_{\mathcal{Y}}(X_i)$, the desired result follows. ■

Based on Lemma 5.1, for uniform data we can, without loss of generality, assume $T(\mathcal{Y}_3)$ to be the basic triangle T_b .

5.1 Transformation of T_b to T_e

There are also transformations that preserve uniformity of the random variable, but not similarity of the triangles. We only describe the transformation that maps $T(\mathcal{Y}_3)$ to the standard equilateral triangle, $T_e = T((0,0), (1,0), (1/2, \sqrt{3}/2))$ for exploiting the symmetry in calculations using T_e .

Let $\phi_e : (x, y) \rightarrow (u, v)$, where $u(x, y) = x + \frac{1-2c_1}{\sqrt{3}}y$ and $v(x, y) = \frac{\sqrt{3}}{2c_2}y$. Then y_1 is mapped to $(0,0)$, y_2 is mapped to $(1,0)$, and y_3 is mapped to $(1/2, \sqrt{3}/2)$. See also Figure 4. Note that the inverse transformation is $\phi_e^{-1}(u, v) = (x(u, v), y(u, v))$ where $x(u, v) = u - \frac{(1-2c_1)}{\sqrt{3}}v$ and $y(u, v) = \frac{2c_2}{\sqrt{3}}v$. Then the Jacobian is given by

$$J(x, y) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & \frac{2c_1-1}{\sqrt{3}} \\ 0 & \frac{2c_2}{\sqrt{3}} \end{vmatrix} = \frac{2c_2}{\sqrt{3}}.$$

So $f_{U,V}(u, v) = f_{X,Y}(\phi_e^{-1}(u, v))|J| = \frac{4}{\sqrt{3}}\mathbf{I}((u, v) \in T_e)$. Hence uniformity is preserved.

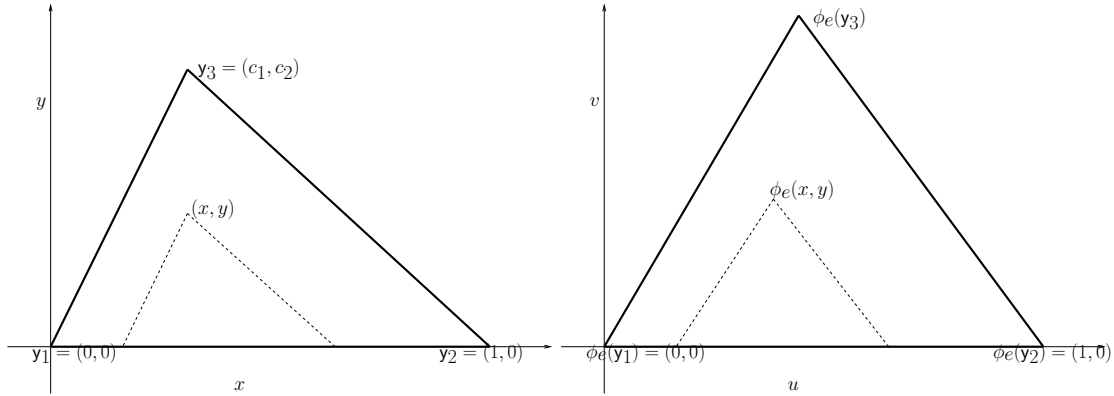


Figure 4: The description of $\phi_e(x, y)$ for $(x, y) \in T_b$ (left) and the equilateral triangle $\phi_e(T_b) = T_e$ (right).

Theorem 5.2. *The arc probability $\mu(N_{\mathcal{Y}})$ of the PCD based on $N_{\mathcal{Y}}$ for uniform data on T_b is geometry invariant iff $A(\phi_e(N_{\mathcal{Y}}(x))) = A(N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x)))$ for all $x \in T_b$.*

Proof: By Lemma 5.1, the PCD based on $N_{\mathcal{Y}}$ for uniform data on $T(\mathcal{Y}_3)$ is rigid-motion and scale invariant. So $T(\mathcal{Y}_3)$ can be transformed to T_b preserving the uniformity of the data and the arc probability for the associated PCD. For uniform data, the set probabilities are calculated as the ratio of the area of the set to the total area. Suppose the arc probability is geometry invariant. Then $\mu(N_{\mathcal{Y}}) = P(X \in N_{\mathcal{Y}}(x)) = P(\phi_e(X) \in N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x)))$. But $P(X \in N_{\mathcal{Y}}(x)) = A(N_{\mathcal{Y}}(x))/A(T_b)$ and $P(\phi_e(X) \in N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x))) = A(N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x)))/A(T_e)$. Moreover $A(N_{\mathcal{Y}}(x))/A(T_b) = A(\phi_e(N_{\mathcal{Y}}(x)))/A(\phi_e(T_b)) = A(\phi_e(N_{\mathcal{Y}}(x)))/A(T_e)$ since the Jacobian cancels out and $\phi_e(T_b) = T_e$. Hence $A(N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x)))/A(T_e) = A(\phi_e(N_{\mathcal{Y}}(x)))/A(T_e)$ implies $A(\phi_e(N_{\mathcal{Y}}(x))) = A(N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x)))$ for all $x \in T_b$. The converse can be proved similarly. ■

Corollary 5.3. *If $\phi_e(N_{\mathcal{Y}}(x)) = N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x))$ for all $x \in T_b$, then the arc probability $\mu(N_{\mathcal{Y}})$ of the PCD based on $N_{\mathcal{Y}}$ for uniform data on T_b is geometry invariant.*

Proof: Let $x \in T_b$. Then $\phi_e(N_{\mathcal{Y}}(x)) = N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x))$ implies $A(\phi_e(N_{\mathcal{Y}}(x))) = A(N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x)))$. Hence the result follows by Theorem 5.2. ■

Definition 5.4. The M -edge regions are said to be geometry invariant if $\phi_e(R_M(e_i)) = R_{\phi_e(M)}(\phi_e(e_i))$ for $i = 1, 2, 3$. The M -vertex regions are said to be geometry invariant if $\phi_e(R_M(y_i)) = R_{\phi_e(M)}(\phi_e(y_i))$ for $i = 1, 2, 3$. □

Corollary 5.5. *Suppose $N_{\mathcal{Y}}$ is based on geometry invariant edge or vertex regions. If the proximity regions are based on boundary of $T(\mathcal{Y}_3)$ and parallel lines to edges, then geometry invariance of the arc probability for uniform data follows.*

Proof: Such proximity maps with geometry invariant edge or vertex regions, satisfy $\phi_e(N_{\mathcal{Y}}(x)) = N_{\phi_e(\mathcal{Y}_3)}(\phi_e(x))$. Hence the desired result follows by Corollary 5.3. ■

Corollary 5.6. *If the edge or vertex regions are based on specific angles in T_b in the sense that their vertices have specific angular values, then these regions are not geometry invariant. Similarly if the proximity regions are based on specific angles in T_b then they are not geometry invariant either.*

Proof: The transformation ϕ_e clearly does not preserve the angles in T_b . Hence the regions dependent on (inner) angles of T_b fail to be preserved. ■

6 Triangle Centers

The PCDs will be defined using the vertex and edge regions, which will be constructed using a point, preferably, in the interior of the triangle, e.g., a *triangle center*. Let $\mathcal{Y}_3 = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$ be non-collinear and $T(\mathcal{Y}_3)$ be the corresponding triangle. The *trilinear coordinates* of a point P with respect to $T(\mathcal{Y}_3)$ are an ordered triple of numbers, which are proportional to the distances from P to the edges. Trilinear coordinates are denoted as $(\alpha : \beta : \gamma)$ and also are known as *homogeneous coordinates* or *trilinears*. Trilinear coordinates were introduced by Plücker in 1835 (see Weisstein (2008)). The triplet of trilinear coordinates obtained by multiplying a given triplet by any positive constant k describes the same point; i.e., $(\alpha : \beta : \gamma) = (k\alpha : k\beta : k\gamma)$, for any $k > 0$. By convention, the three vertices y_1 , y_2 , and y_3 of $T(\mathcal{Y}_3)$ are commonly written as $(1 : 0 : 0)$, $(0 : 1 : 0)$, and $(0 : 0 : 1)$, respectively (see Weisstein (2008)).

Definition 6.1. A *triangle center* is a point whose trilinear coordinates are defined in terms of the edge lengths and (inner) angles of a triangle. The function giving the coordinates $(\alpha : \beta : \gamma)$ is called the *triangle center function*. □

Kimberling (2008) enumerates 360 triangle centers, among which four have been widely known since the ancient times; namely, *circumcenter* (CC), *incenter* (IC), *center of mass* or *centroid* (CM), and *orthocenter* (OC). The point where the center is located in $T(\mathcal{Y}_3)$ will be labeled accordingly; e.g., M_{CC} will denote the circumcenter of $T(\mathcal{Y}_3)$.

The *circumcircle* is a triangle's circumscribed circle; i.e., the unique circle that passes through each of the triangle's three vertices, y_1, y_2, y_3 . The center of the circumcircle is called the *circumcenter*, denoted as M_{CC} , and the circle's radius is called the *circumradius*, denoted as r_{cc} . By construction, the distances from circumcenter to the vertices are equal (to r_{cc}). Furthermore, the triangle's three edge bisectors perpendicular to edges e_i at M_i for $i \in \{1, 2, 3\}$ intersect at M_{CC} . See Figure 5. The trilinear coordinates of M_{CC} are $(\cos \theta_1 : \cos \theta_2 : \cos \theta_3)$ where θ_i is the inner angle of $T(\mathcal{Y}_3)$ at vertex y_i for $i \in \{1, 2, 3\}$ and the trilinears for M_{CC} can also be written as $(r_{cc} \cos \theta_1 : r_{cc} \cos \theta_2 : r_{cc} \cos \theta_3)$.

The circumcenter of a triangle is in the interior, at the midpoint of the hypotenuse, or in the exterior of the triangle, if the triangle is acute, right, or obtuse, respectively. See Figure 5 where an acute and an obtuse triangle are depicted. Using the pdf of an arbitrary angle of a triangle T_j from Poisson Delaunay triangulation \mathcal{D}_P (Mardia et al. (1977)), we see that,

$$P(T_j \text{ is a right triangle}) = P(\theta = \pi/2) = 0,$$

hence $P(M_{CC} \text{ is the midpoint of the hypotenuse}) = 0$. Furthermore,

$$\begin{aligned} P(T_j \text{ is an obtuse triangle}) &= P(M_{CC} \notin T_j) = P(\theta_{\max} > \pi/2) = \int_{\pi/2}^{\pi} f_3(x) dx \\ &= \frac{(3 f_S(\sqrt{2}\pi) - f_C(\sqrt{2}\pi) - 3 f_S(\sqrt{\pi/2}) + f_C(\sqrt{\pi/2}))}{\sqrt{2}\pi} \\ &\approx .03726 \end{aligned}$$

where

$$\begin{aligned} f_3(x) &= \left[\frac{2}{\pi} (3x(\sin 2x) - \cos 2x + \cos 4x - \pi \sin 2x) \right] \mathbf{I}(\pi/3 < x < \pi/2) + \\ &\quad \left[\frac{1}{\pi} (4\pi(\cos x)(\sin x) + 3 \sin^2 x - \cos^2 x - 4x(\cos x)(\sin x) + 1) \right] \mathbf{I}(\pi/2 < x < \pi) \end{aligned}$$

is the pdf of the maximum angle, $f_C(x) = \int_0^x \cos(\pi t^2/2) dt$, and $f_S(x) = \int_0^x \sin(\pi t^2/2) dt$ are the Fresnel cosine and sine functions, respectively. The coordinates of M_{CC} in the basic triangle T_b are $\left(\frac{1}{2}, \frac{c_1^2 - c_1 + c_2^2}{2c_2}\right)$.

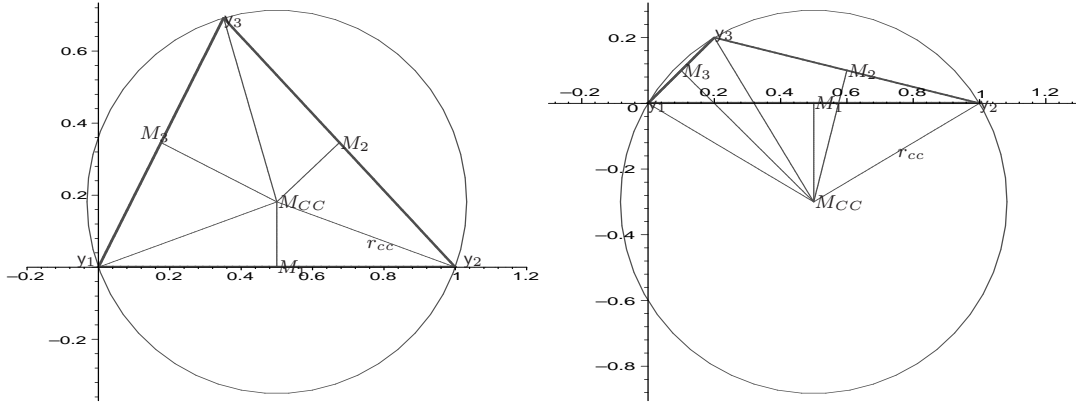


Figure 5: The circumcenter, circumcircle, and circumradius of an acute triangle (left), an obtuse triangle (right).

The *incircle* is the inscribed circle of a triangle, i.e., the unique circle that is tangent to the edges e_i at P_i for $i \in \{1, 2, 3\}$. The center of the incircle is called the *incenter*, denoted as M_I , and the radius of the incircle is called the *inradius*, denoted as r_{ic} . Incenter has trilinear coordinates $(1 : 1 : 1)$. The incenter is the point where the triangle's inner angle bisectors meet. See Figure 6 (left).

The coordinates of M_I for the basic triangle T_b are (x_{ic}, y_{ic}) , where

$$x_{ic} = \frac{c_1 - \sqrt{c_1^2 + c_2^2}}{1 + \sqrt{c_1^2 + c_2^2} + \sqrt{(1 - c_1)^2 + c_2^2}}, \quad y_{ic} = \frac{c_2}{1 + \sqrt{c_1^2 + c_2^2} + \sqrt{(1 - c_1)^2 + c_2^2}}.$$

Note that, M_{CC} and M_I do not necessarily concur. The distance between M_{CC} and M_I is $d(M_{CC}, M_I) = \sqrt{r_{cc}(r_{cc} - 2r_{ic})}$. Unlike the circumcenter, the incenter is guaranteed to be inside the triangle.

The *median line* of a triangle is the line from one of its vertices to the midpoint of the opposite edge. The three median lines of any triangle intersect at the triangle's *centroid*, denoted as M_C . The centroid is the *center of mass* of the vertices of a triangle. Since M_C is also the intersection of the triangle's three median lines, it is sometimes called the *median point*. It has trilinear coordinates $(1/|e_1| : 1/|e_2| : 1/|e_3|)$ or $(\csc \theta_1 : \csc \theta_2 : \csc \theta_3)$ where e_i denotes the edge opposite to the vertex y_i for $i \in \{1, 2, 3\}$. The centroid is also guaranteed to be in the interior of the triangle. See Figure 6 (right). The coordinates of M_C for the basic triangle are $((1 + c_1)/3, c_2/3)$.

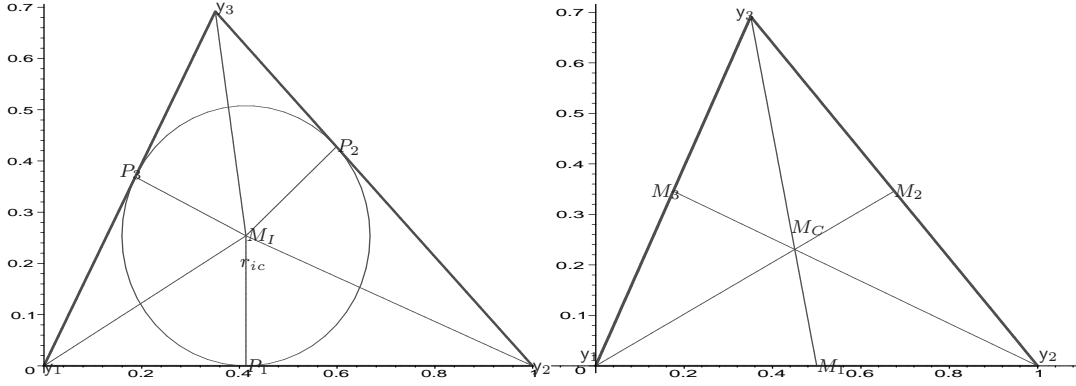


Figure 6: The incircle, incenter, inradius of a triangle (left), and the centroid or center of mass of a triangle (right).

The intersection of the three altitudes of a triangle is called the *orthocenter*, M_O , which has trilinear coordinates $(\cos \theta_2 \cos \theta_3 : \cos \theta_1 \cos \theta_3 : \cos \theta_1 \cos \theta_2)$. The orthocenter of a triangle is in the interior, at vertex y_3 , or in the exterior of the basic triangle, T_b , if T_b is acute, right, or obtuse, respectively. The functional form of M_O in the basic triangle is $(c_1, c_1(1 - c_1)/c_2)$.

Note that in an equilateral triangle, $M_I = M_{CC} = M_O = M_C$ (i.e., all four centers we have described coincide).

7 Vertex and Edge Regions

The new proximity maps will be based on the Delaunay cell \mathcal{T}_j that contains x . The region $N_{\mathcal{Y}}(x)$ will also depend on the location of x in \mathcal{T}_j with respect to the vertices or faces (edges in \mathbb{R}^2) of \mathcal{T}_j . Hence for $N_{\mathcal{Y}}(x)$ to be well-defined, the vertex or face of \mathcal{T}_j associated with x should be uniquely determined. This will give rise to two new concepts: *vertex regions* and *face regions* (edge regions in \mathbb{R}^2).

7.1 Vertex Regions

Let $\mathcal{Y}_3 = \{y_1, y_2, y_3\}$ be three non-collinear points in \mathbb{R}^2 and $T(\mathcal{Y}_3) = T(y_1, y_2, y_3)$ be the triangle with vertices \mathcal{Y}_3 . Then for $x \in T(\mathcal{Y}_3)$, $N_S(x) = B(x, r(x))$ where $r(x) = \min_{y \in \mathcal{Y}_3} d(x, y)$. That is, $r(x) = d(x, y_i)$ iff $x \in \mathcal{V}_C(y_i) \cap T(\mathcal{Y}_3)$ for $i \in \{1, 2, 3\}$, where $\mathcal{V}_C(y_i)$ is the Voronoi cell generated by y_i in the Voronoi tessellation based on \mathcal{Y}_3 . Notice that these cells partition the triangle $T(\mathcal{Y}_3)$ and each $\mathcal{V}_C(y_i) \cap T(\mathcal{Y}_3)$ is adjacent only to vertex y_i and their intersection is the point M which is equidistant to the vertices, so M is in fact the circumcenter, M_{CC} , of $T(\mathcal{Y}_3)$. See Figure 7. To define new proximity regions based on some sort

of distance or dissimilarity relative to the vertices \mathcal{V}_3 , we associate each point in $T(\mathcal{V}_3)$ to a vertex of $T(\mathcal{V}_3)$ as in the spherical case. This gives rise to the concept of *vertex regions*. Note that $N_S(x)$ is constructed using the vertex region based on the closest vertex, $\operatorname{argmin}_{y \in \mathcal{V}_3} d(x, y)$. If two vertices were equidistant from x (i.e., $\operatorname{argmin}_{y \in \mathcal{V}_3} d(x, y)$ were not unique), x is arbitrarily assigned to a vertex region. In fact, for N_S , by construction, it would not matter which vertex to pick when the vertices are equidistant to x , the region $N_S(x)$ will be the same.

Definition 7.1. The connected regions that partition the triangle, $T(\mathcal{V}_3)$, (in the sense that the intersections of the regions have zero \mathbb{R}^2 -Lebesgue measure) such that each region has one and only one vertex of $T(\mathcal{V}_3)$ on its boundary are called *vertex regions*. \square

This definition implies that there are three vertex regions. In fact, the vertex regions can be described starting with a point $M \in \mathbb{R}^2 \setminus \mathcal{V}_3$. Join the point M to a point on each edge by a curve such that the resultant regions satisfy the above definition. Such regions are called *M-vertex regions* and we denote the vertex region associated with vertex y as $R_M(y)$ for $y \in \mathcal{V}_3$. In particular, one can use a *center* of the triangle $T(\mathcal{V}_3)$ as the starting point M for vertex regions. See the discussion of triangle centers in Section 6. The points in $R_M(y)$ can be thought as being “closer” to y than to the other vertices.

It is reasonable to require that the area of the region $R_M(y)$ gets larger as $d(M, y)$ increases. Usually the curves will be taken to be lines or even the orthogonal projections to the edges. But these lines do not necessarily yield three vertex regions for M in the exterior of $T(\mathcal{V}_3)$. Unless stated otherwise, *M-vertex regions* will refer to regions constructed by joining M to the edges with *straight line segments*.

7.1.1 M-Vertex Regions

For $M \in T(\mathcal{V}_3)^o$, *M-vertex regions* are defined by two ways:

(I) Geometrically, one can construct *M-vertex regions* by drawing the **orthogonal projections to the edges**, denoted as $R_M^\perp(y)$. For instance see Figure 7 with $M = M_{CC}$.

The functional forms of $R_M^\perp(y)$ for $M = (m_1, m_2)$ in the basic triangle are:

$$\begin{aligned} R_M^\perp(y_1) &= \{(x, y) \in T(\mathcal{V}_3) : x \leq m_1; y \leq m_2 - (x - m_1)c_1/c_2\}, \\ R_M^\perp(y_2) &= \{(x, y) \in T(\mathcal{V}_3) : x \geq m_1; y \leq m_2 + (1 - c_1)(x - m_1)/c_2\}, \text{ and} \\ R_M^\perp(y_3) &= \{(x, y) \in T(\mathcal{V}_3) : y \geq m_2 - c_1(x - m_1)/c_2; y \geq m_2 + (1 - c_1)(x - m_1)/c_2\}. \end{aligned}$$

However, the orthogonal projections from M to the edges does not necessarily fall on the boundary of $T(\mathcal{V}_3)$. For example, letting P_2^M be the orthogonal projection of M to edge e_2 , it is easy to see that P_2^M might fall outside $T(\mathcal{V}_3)$ which contradicts the definition of vertex regions. In fact $P_2^M \in e_2$ iff $\frac{c_2(m_2 c_2 + c_1 m_1)}{c_1^2 + c_2^2} \leq c_2$ iff $c_2(c_2 - m_2) + c_1(c_1 - m_1) \geq 0$.

(II) One can also construct *M-vertex regions* with $M \in T(\mathcal{V}_3)^o$ by **using the extensions of the line segments joining y to M** for all $y \in \mathcal{V}_3$. See Figure 9 with $M = M_C$. The functional forms of $R_M(y_i)$ for $i \in \{1, 2, 3\}$ with $M = (m_1, m_2)$ and $m_1 > c_1$ in the basic triangle, T_b , are given by

$$\begin{aligned} R_M(y_1) &= \left\{ (x, y) \in T(\mathcal{V}_3) : y \leq \frac{m_2(x-1)}{m_1-1}; y \leq \frac{m_2(c_1-x) + c_2(x-m_1)}{c_1-m_1} \right\}, \\ R_M(y_2) &= \left\{ (x, y) \in T(\mathcal{V}_3) : y \leq \frac{m_2 x}{m_1}; y \geq \frac{m_2(c_1-x) + c_2(x-m_1)}{c_1-m_1} \right\}, \text{ and} \\ R_M(y_3) &= \left\{ (x, y) \in T(\mathcal{V}_3) : y \geq \frac{m_2 x}{m_1}; y \geq \frac{m_2(x-1)}{m_1-1} \right\}. \end{aligned}$$

For $m_1 < c_1$, $R_M(y_i)$ for $i \in \{1, 2, 3\}$ are defined similarly.

If x falls on the boundary of two M -vertex regions, then x is arbitrarily assigned to one of the M -vertex regions.

To distinguish between these two types, the vertex regions constructed by using orthogonal projections are denoted as $R_M^\perp(y)$ and the vertex regions constructed by using the lines joining vertices to M are denoted as $R_M(y)$. By definition, $R_{OC}(y)$ and $R_{OC}^\perp(y)$ are identical. But, for $M = M_C, M_{CC}, M_I$, $R_M(y)$ can have both versions, so the above distinction is necessary for them.

7.1.2 CC -Vertex Regions

The region $\mathcal{V}_C(y) \cap T(\mathcal{Y}_3)$ is a special type of vertex regions, which can also be obtained geometrically by starting at M_{CC} and drawing the orthogonal projections to the edges. Hence these regions are called *CC-vertex regions*. One can also construct CC -vertex regions by drawing the perpendicular (mid)edge bisectors or by finding the circumcenter and drawing the orthogonal projections to the edges. See Figure 7, where M_i are the midpoints of the edges.

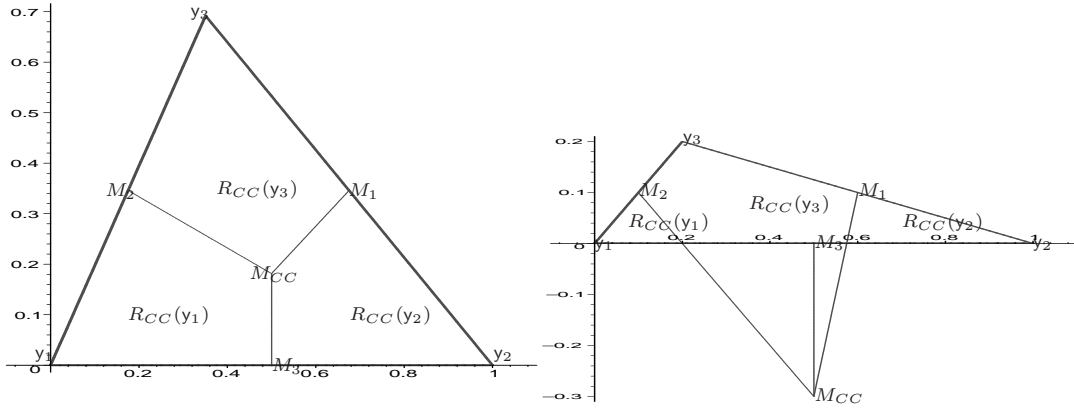


Figure 7: The CC -vertex regions in an acute triangle (left) and in an obtuse triangle (right).

The functional forms of $R_{CC}(y_i)$ for $i \in \{1, 2, 3\}$, in the basic triangle $T_b = ((0, 0), (1, 0), (c_1, c_2))$ (see Section 5) are given by

$$\begin{aligned} R_{CC}(y_1) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : x \leq \frac{1}{2}; y \leq \frac{c_1^2 + c_2^2 - 2c_1x}{2c_2} \right\}, \\ R_{CC}(y_2) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : x \geq \frac{1}{2}; y \leq \frac{c_1^2 + c_2^2 - 1 + 2(1 + c_1)x}{2c_2} \right\}, \\ R_{CC}(y_3) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \geq \frac{c_1^2 + c_2^2 - 2c_1x}{2c_2}; y \geq \frac{c_1^2 + c_2^2 - 1 + 2(1 + c_1)x}{2c_2} \right\}. \end{aligned}$$

One can also define CC -vertex regions by using the line segments which join M_{CC} to edge e_i and are extensions of the lines joining M_{CC} to the vertex v_i for $i \in \{1, 2, 3\}$, but this definition only works for acute triangles, since $M_{CC} \notin T(\mathcal{Y}_3)^\circ$ for non-acute triangles.

7.1.3 CM -Vertex Regions

The motivation behind CM -vertex regions is that unlike the circumcenter, center of mass is guaranteed to be inside the triangle. We define the CM -vertex regions using the median lines and denote the regions as

$R_{CM}(y_i)$ for $i \in \{1, 2, 3\}$ (see Figure 9). However, the method with orthogonal projections of M_C to the edges does not always work. Let P_i^{CM} be the point at which orthogonal projection of M_C on e_i crosses e_i for $i \in \{1, 2, 3\}$. Then, P_2^{CM} might fall outside $T(\mathcal{Y}_3)$ in which case $R_{CM}^\perp(y_1)$ is adjacent to two vertices y_1 and y_3 , while $R_{CM}^\perp(y_3)$ is not adjacent to any of the vertices. Hence the definition of the vertex regions is violated. In fact $P_2^{CM} \in e_2$ iff $\frac{c_2(c_2^2 + c_1^2 + c_1)}{3(c_1^2 + c_2^2)} \leq c_2$ iff $2c_1^2 + 2c_2^2 - c_1 \geq 0$. See Figure 8 for the domain of (c_1, c_2) in T_b for $P_2^{CM} \in e_2$.

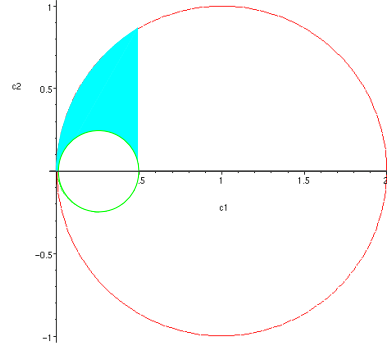


Figure 8: Depicted is the domain of (c_1, c_2) where $P_2^{CM} \in e_2$.

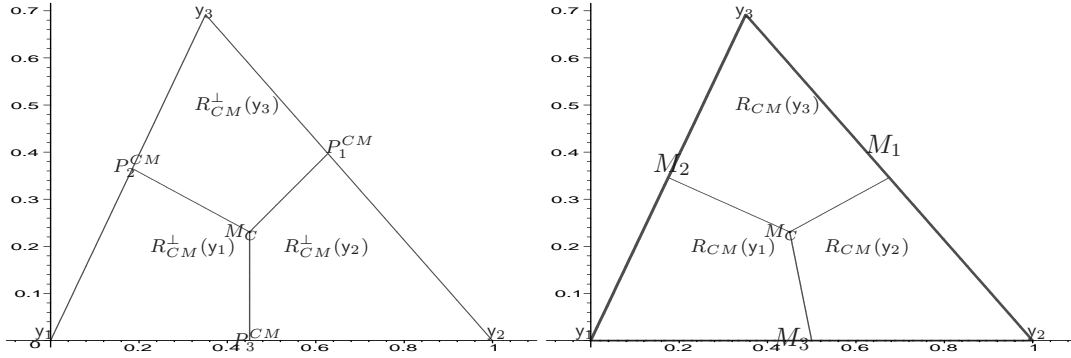


Figure 9: The CM -vertex regions with orthogonal projections (left) and with median lines (right).

The functional forms of $R_{CM}(y_i)$ for $i \in \{1, 2, 3\}$ in the basic triangle T_b are given by

$$\begin{aligned} R_{CM}(y_1) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \leq \frac{c_2(2x-1)}{2c_1-1}; y \leq \frac{c_2(x-1)}{c_1-2} \right\}, \\ R_{CM}(y_2) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \leq \frac{c_2(2x-1)}{2c_1-1}; y \leq \frac{c_2 x}{1+c_1} \right\}, \\ R_{CM}(y_3) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \geq \frac{c_2(x-1)}{c_1-2}; y \geq \frac{c_2 x}{1+c_1} \right\}. \end{aligned}$$

7.1.4 IC-Vertex Regions

One can also define the incenter vertex regions by using the inner angle bisectors. With orthogonal projections, *IC*-vertex regions are bounded by the edges of $T(\mathcal{Y}_3)$ and the inradii crossing the tangential points of the incircle on the edges. These three regions are denoted as $R_{IC}^\perp(y)$ for $y \in \mathcal{Y}_3$. With the inner angle bisectors, the incenter M_I is used and the parts of the inner angle bisectors that join M_I to the edges. These vertex regions are denoted as $R_{IC}(y)$. See Figure 10 for both versions of the vertex regions.

Note that one might also use the orthocenter, M_O , to define the vertex regions. However, for non-acute triangles *OC*-vertex regions cannot naturally be defined.

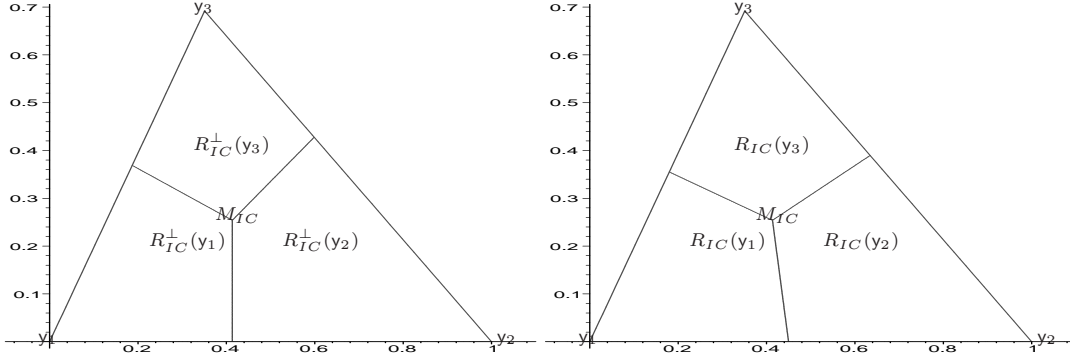


Figure 10: The *IC*-vertex regions with orthogonal projections (left) and with extension of the line segments joining the vertices to M_{IC} (right).

7.2 Edge Regions

The spherical proximity region seen earlier is constructed by using the vertex region based on the closest vertex, $\operatorname{argmin}_{y \in \mathcal{Y}_3} d(x, y)$. One can also use the closest edge, $\operatorname{argmin}_{i \in \{1, 2, 3\}} d(x, e_i)$, in defining a proximity region, which suggests the concept of *edge regions*.

While using the edge $\operatorname{argmin}_{i \in \{1, 2, 3\}} d(x, e_i)$, the triangle is again partitioned into three regions whose intersection is some point M with Euclidean distance to the edges $d(M, e_1) = d(M, e_2) = d(M, e_3)$, so M is in fact the incenter of $T(\mathcal{Y}_3)$ and $d(M, e) = r_{ic}$ is the inradius (see Section 6 for incenter and inradius).

Definition 7.2. The connected regions that partition the triangle, $T(\mathcal{Y}_3)$, in such a way that each region has one and only one edge of $T(\mathcal{Y}_3)$ on its boundary, are called *edge regions*. \square

This definition implies that there are exactly three edge regions which intersect at only one point, M in $T(\mathcal{Y}_3)^o$, the interior of $T(\mathcal{Y}_3)$. In fact, one can describe the edge regions starting with M . Join the point M to the vertices by curves such that the resultant regions satisfy the above definition. Such regions are called *M-edge regions* and the edge region for edge e is denoted as $R_M(e)$ for $e \in \{e_1, e_2, e_3\}$. Unless stated otherwise, *M*-edge regions will refer to the regions constructed by joining M to the vertices by straight lines. In particular, one can use a *center* of $T(\mathcal{Y}_3)$ for the starting point M . One can also consider the points in $R_M(e)$ to be “closer” to e than to the other edges. Furthermore, it is reasonable to require that the area of the region $R_M(e)$ get larger as $d(M, e)$ increases. Moreover, in higher dimensions, the corresponding regions are called “face regions”.

The functional forms of $R_M(e_i)$ for $i \in \{1, 2, 3\}$, for $M = (m_1, m_2) \in T(\mathcal{Y}_3)^o$ and $m_1 > c_1$ in the basic

triangle are given by

$$\begin{aligned} R_M(e_1) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \geq \frac{m_2(x-1)}{m_1-1}; y \geq \frac{c_2(x-m_1)-m_2(x-c_1)}{c_1-m_1} \right\}, \\ R_M(e_2) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \geq \frac{m_2 x}{m_1}; y \leq \frac{c_2(x-m_1)-m_2(x-c_1)}{c_1-m_1} \right\}, \text{ and} \\ R_M(e_3) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : \frac{m_2 x}{m_1}; y \leq \frac{m_2(x-1)}{m_1-1} \right\}. \end{aligned}$$

If x falls on the boundary of two M -edge regions, then x is arbitrarily assigned to one of the M -edge regions.

The center of mass edge regions (CM -edge regions) are described in detail, as we will use them in defining a new class of proximity maps.

7.2.1 CM -Edge Regions

One can divide $T(\mathcal{Y}_3)$ into three regions by using the median lines which intersect at the centroid, or equivalently, joining the centroid M_C to the vertices by straight lines will yield the CM -edge regions. Let $R_{CM}(e)$ be the region for edge $e \in \{e_1, e_2, e_3\}$. See Figure 11 (left).

The functional forms of $R_{CM}(e_i)$ for $i \in \{1, 2, 3\}$, in the basic triangle, T_b , are given by

$$\begin{aligned} R_{CM}(e_1) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \leq \frac{c_2(1-2x)}{1-2c_1}; y \geq \frac{c_2(1-x)}{2-c_1} \right\}, \\ R_{CM}(e_2) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \geq \frac{c_2(1-2x)}{1-2c_1}; y \geq \frac{c_2 x}{1+c_1} \right\}, \\ R_{CM}(e_3) &= \left\{ (x, y) \in T(\mathcal{Y}_3) : y \leq \frac{c_2 x}{1+c_1}; y \leq \frac{c_2(1-x)}{2-c_1} \right\}. \end{aligned}$$

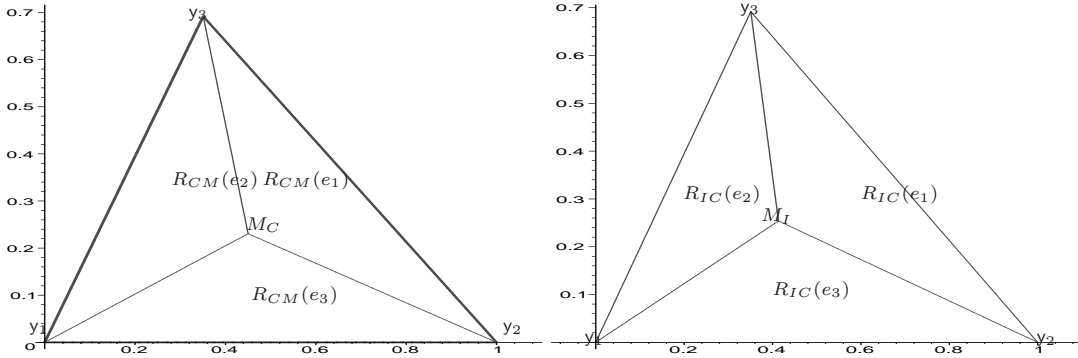


Figure 11: CM -edge regions $R_{CM}(e_i)$ (right) and $R_{IC}(e_i)$ (left) for $i \in \{1, 2, 3\}$.

Remark 7.3. One can also divide $T(\mathcal{Y}_3)$ into three regions by using the inner angle bisectors which intersect at the incenter, yielding the incenter edge regions (IC -edge regions). Let $R_{IC}(e)$ be the region for edge $e \in \{e_1, e_2, e_3\}$. Notice that the closest edge to any point in $R_{IC}(e)$ is edge e , i.e., $x \in R_{IC}(e)$ iff $\operatorname{argmin}_{u \in \{e_1, e_2, e_3\}} d(x, u) = e$. If two edges are equidistant from x , x is arbitrarily assigned to an edge region. See Figure 11 (right). \square

Remark 7.4. In \mathbb{R} , one can view the end points of $[0, 1]$, $\{0, 1\}$, as vertices or edges. So $[0, 1/2]$ and $[1/2, 1]$ can be viewed as either vertex regions or edge regions. \square

8 Proximity Regions in Delaunay Tessellations

Let $\mathcal{Y}_m = \{y_1, \dots, y_m\}$ be m points in general position in \mathbb{R}^d and \mathcal{T}_j be the j^{th} Delaunay cell for $j = 1, \dots, J$. Let also that \mathcal{X}_n be a random sample from F with support $\mathcal{S}(F) \subseteq \mathcal{C}_H(\mathcal{Y}_m)$. That is, $\Omega = \mathcal{C}_H(\mathcal{Y}_m)$ and the $\Omega_j = \mathcal{T}_j$ with μ being the Lebesgue measure. Then the appealing properties for proximity regions in Section 4 become:

- P1** $N(x)$ is well defined for all $x \in \mathcal{C}_H(\mathcal{Y}_m)$.
- P2** $x \in N(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y}_m)$.
- P3** x is at the *center* of $N(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y}_m)$.
- P4** For $x \in \mathcal{T}_j \subseteq \mathcal{C}_H(\mathcal{Y}_m)$, $N(x)$ and \mathcal{T}_j are of the *same type*; they are both $(d+1)$ -simplices.
- P5** For $x \in \mathcal{T}_j \subseteq \mathcal{C}_H(\mathcal{Y}_m)$, $N(x)$ mimics the shape of \mathcal{T}_j ; i.e., it is *similar* to \mathcal{T}_j .
- P6** Conditional on $X \in \mathcal{T}_j$, $N(X)$ is a proper subset of \mathcal{T}_j a.s.
- P7** For $x \in \mathcal{T}_j$ and $y \in \mathcal{T}_k$ with $j \neq k$, $N(x)$ and $N(y)$ are disjoint a.s.
- P8** The size of $N(x)$ is continuous in x ; that is, for each $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that $|\mu(N(y)) - \mu(N(x))| < \delta(\varepsilon)$ whenever $\|y - x\| < \varepsilon$.
- P9** The arc probability $\mu(N_S)$ does not depend on the support region for uniform data in \mathbb{R}^d .

In particular, for illustrative purposes, we focus on \mathbb{R}^2 , where a Delaunay tessellation is a triangulation, provided that no more than three points of \mathcal{Y}_m are cocircular. Furthermore, for simplicity, let $\mathcal{Y}_3 = \{y_1, y_2, y_3\}$ be three non-collinear points in \mathbb{R}^2 and $T(\mathcal{Y}_3) = T(y_1, y_2, y_3)$ be the triangle with vertices \mathcal{Y}_3 . Let \mathcal{X}_n be a random sample from F with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. The spherical proximity map is the first proximity map defined in literature (see DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a), Priebe et al. (2003b), and DeVinney and Priebe (2006)) where *CC*-vertex regions were implicitly used for points in $\mathcal{C}_H(\mathcal{Y}_m)$. In the following sections, we will describe arc-slice proximity maps $N_{AS}(\cdot)$ and define two families of triangular proximity regions for which **P4** and **P5** will automatically hold.

8.1 Arc-Slice Proximity Maps

Recall that for $N_S(\cdot)$ **P7** is violated, since for any $x \in \mathcal{T}_j \subset \mathbb{R}^d$, $B(x, r(x)) \not\subset \mathcal{T}_j$, which implies that two proximity regions $N_S(x)$ and $N_S(y)$ might overlap for x, y in two distinct cells. Such an overlap of the regions make the distribution of the domination number of the PCD associated with $N_S(\cdot)$, if not impossible, hard to calculate. In order to avoid the overlap of regions $B(x, r(x))$ and $B(y, r(y))$ for x, y in different Delaunay cells, the balls are restricted to the corresponding cells, which leads to *arc-slice proximity regions*, $N_{AS}(x) := \overline{B}(x, r(x)) \cap \mathcal{T}_j$, where $\overline{B}(x, r(x))$ is the closure of the ball $B(x, r(x))$. The closed ball is used in the definition of the arc-slice proximity map for consistency with the other proximity maps that will be defined on Delaunay cells. The arc-slice proximity map $N_{AS}(x)$ is well-defined only in $\mathcal{C}_H(\mathcal{Y}_m)$, provided that \mathcal{Y}_m is in general position and $m \geq d+1$ in \mathbb{R}^d .

By construction, the *CC*-vertex regions are implicitly used, since $x \in R_{CC}(y)$ iff $y = \operatorname{argmin}_{u \in \mathcal{Y}_m} d(x, u)$. To make this dependence explicit, the notation $N_{AS}(\cdot, M_{CC})$ is used. See Figure 12 (top) for $N_{AS}(x, M_{CC})$ for an $x \in R_{CC}(y_2)$. The functional form of $N_{AS}(x, M_{CC})$ for an $x \in R_{CC}(y)$ is given by

$$N_{AS}(x, M_{CC}) := \{z \in T(\mathcal{Y}_3) : d(z, x) \leq r(x) = d(x, y)\}.$$

Notice that, the region $N_{AS}(x, M_{CC})$ is a closed region, unlike $N_S(x) = B(x, r(x))$. The properties **P1**, **P2**, **P7** hold by definition. Notice that $N_{AS}(x, M_{CC}) \subseteq T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3)$ and $N_{AS}(x, M_{CC}) = T(\mathcal{Y}_3)$ iff $x = M_{CC}$, since $\overline{B}(x, r(x)) \supset T(\mathcal{Y}_3)$ only when $x = M_{CC}$. Hence the superset region for arc-slice proximity maps with CC -vertex regions is $\mathcal{R}_S(N_{AS}, M_{CC}) = \{M_{CC}\}$. So **P6** follows. Furthermore, **P8** holds since the area $A(N_{AS}(x, M_{CC}))$ is a continuous function of $r(x) = \min_{y \in \mathcal{Y}_3} d(x, y)$ which is a continuous function of x . **P3**, **P4**, **P5**, and **P9** fail for $N_{AS}(x, M_{CC})$. See Figure 13 for the arcs based on $N_{AS}(x, M_C)$ for a realization of 7 \mathcal{X} points in the one triangle case, and Figure 14 for the arcs for the realization of 77 \mathcal{X} points in the multi-triangle case in Figure 2 (top right).

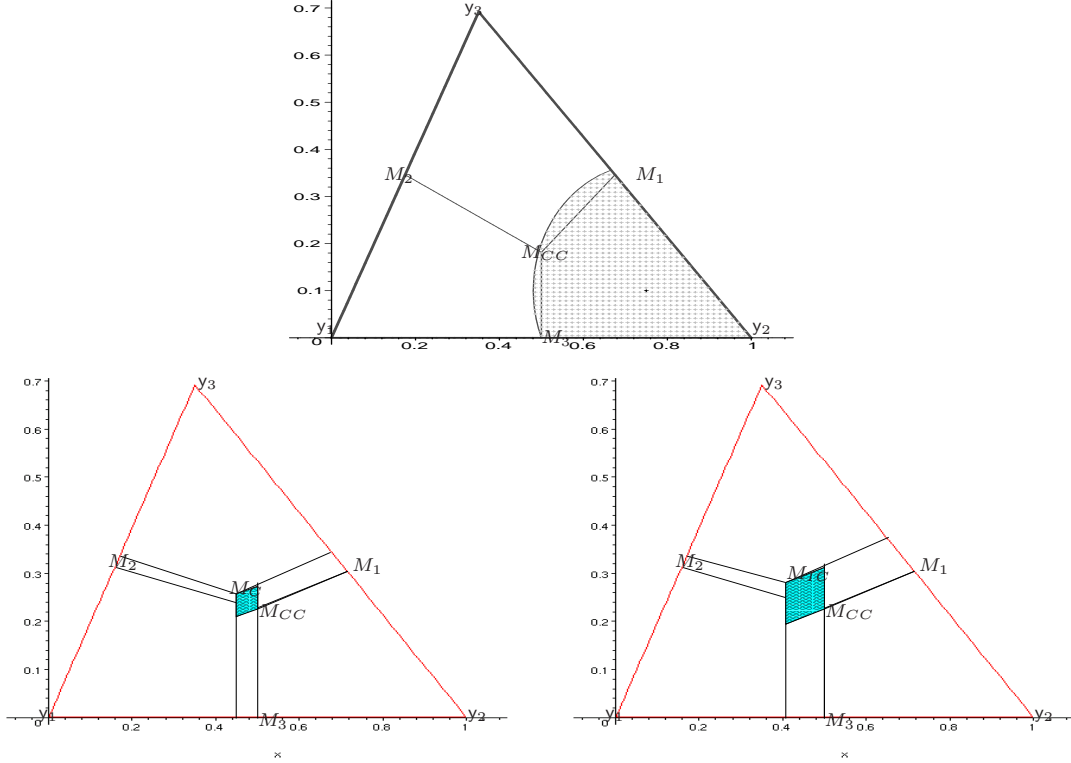


Figure 12: $N_{AS}(x, M_{CC})$ with an $x \in R_{CC}(y_2)$ (top) and the superset regions $\mathcal{R}_S(N_{AS}, M_C)$ (bottom left) $\mathcal{R}_S(N_{AS}, M_{IC})$ in $T(\mathcal{Y}_3)$ (bottom right).

One can define arc-slice proximity regions with any type of M -vertex regions as

$$N_{AS}(x, M) := \overline{B}(x, r(x)) \cap T(\mathcal{Y}_3) \text{ where } r(x) := d(x, y) \text{ for } x \in R_M(y).$$

But for $M \neq M_{CC}$, $N_{AS}(\cdot, M)$ satisfies only **P1**, **P2**, and **P7**. **P6** fails to hold since $\mathcal{R}_S(N_{AS}, M)$ has positive area and **P8** fails since the size of $N_{AS}(x, M)$ is not continuous in x . See, for example, Figure 12 (right) for $\mathcal{R}_S(N_{AS}, M_C)$. In terms of the properties in Section 4, $N_{AS}(\cdot, M_{CC})$ is the most appealing proximity map among the family $\mathcal{N}_{AS} := \{N_{AS}(\cdot, M) : M \in \mathbb{R}^2 \setminus \mathcal{Y}_3\}$.

Moreover, $\Lambda_0(N_{AS}, M) = \mathcal{Y}_3$ for all $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ since $\lambda(N_{AS}(x, M)) = 0$ iff $x \in \mathcal{Y}_3$.

Next, we define *triangular proximity regions*, which, by definition, will satisfy properties **P4** and **P5**. These proximity regions are the building blocks of the PCDs for which more rigorous mathematical analysis — compared to the PCDs based on spherical and arc-slice proximity maps — will be possible.

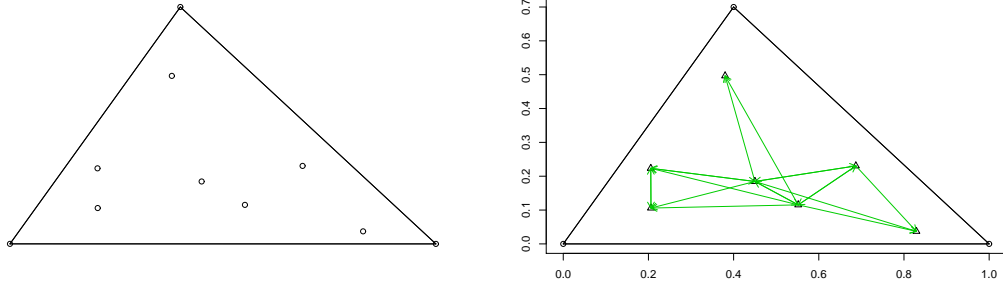


Figure 13: A realization of 7 \mathcal{X} points generated iid $\mathcal{U}(T(\mathcal{Y}_3))$ (left) and the corresponding arcs for $N_{AS}^{r=2}(x, M_C)$ (right).

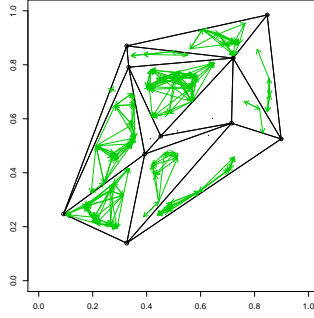


Figure 14: The arcs for arc-slice PCDs with $N_{AS}^{r=2}(x, M_C)$ for the 77 \mathcal{X} points that lie in the $\mathcal{C}_H(\mathcal{Y}_{10})$ (see Figure 2).

8.2 Proportional-Edge Proximity Maps

The first type of triangular proximity map introduced is the proportional-edge proximity map. For this proximity map, the asymptotic distribution of domination number and the relative density of the corresponding PCD will have mathematical tractability (Ceyhan and Priebe (2005), Ceyhan et al. (2006) and Ceyhan and Priebe (2007)).

For $r \in [1, \infty]$, define $N_{PE}^r(\cdot, M) := N(\cdot, M; r, \mathcal{Y}_3)$ to be the *proportional-edge proximity map* with M -vertex regions as follows (see also Figure 15 with $M = M_C$ and $r = 2$). For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $v(x) \in \mathcal{Y}_3$ be the vertex whose region contains x ; i.e., $x \in R_M(v(x))$. If x falls on the boundary of two M -vertex regions, $v(x)$ arbitrarily assigned. Let $e(x)$ be the edge of $T(\mathcal{Y}_3)$ opposite $v(x)$. Let $\ell(v(x), x)$ be the line parallel to $e(x)$ through x . Let $d(v(x), \ell(v(x), x))$ be the Euclidean (perpendicular) distance from $v(x)$ to $\ell(v(x), x)$. For $r \in [1, \infty)$, let $\ell_r(v(x), x)$ be the line parallel to $e(x)$ such that

$$\begin{aligned} d(v(x), \ell_r(v(x), x)) &= r d(v(x), \ell(v(x), x)) \\ \text{and} \\ d(\ell(v(x), x), \ell_r(v(x), x)) &< d(v(x), \ell_r(v(x), x)). \end{aligned}$$

Let $T_r(x)$ be the triangle similar to and with the same orientation as $T(\mathcal{Y}_3)$ having $v(x)$ as a vertex and

$\ell_r(v(x), x)$ as the opposite edge. Then the r -factor proportional-edge proximity region $N_{PE}^r(x, M)$ is defined to be $T_r(x) \cap T(\mathcal{Y}_3)$. Notice that $\ell(v(x), x)$ divides the edges of $T_r(x)$ (other than $\ell_r(v(x), x)$) proportionally with the factor r . Hence the name *proportional edge proximity region* and the notation $N_{PE}^r(\cdot, M)$.

Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M)$. Furthermore, $\lim_{r \rightarrow \infty} N_{PE}^r(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, so $N_{PE}^\infty(x, M) := T(\mathcal{Y}_3)$ for all such x . For $x \in \mathcal{Y}_3$, $N_{PE}^r(x, M) := \{x\}$ for all $r \in [1, \infty]$. See Figure 16 for the arcs based on $N_{PE}^{r=2}(x, M_C)$ in the one triangle and the multi-triangle cases.

Notice that $X_i \stackrel{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional pdf f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of $N_{PE}^r - X$ falls on the boundary of two vertex regions — occurs with probability zero. Note that for such an F , $N_{PE}^r(X, M)$ is a triangle a.s.

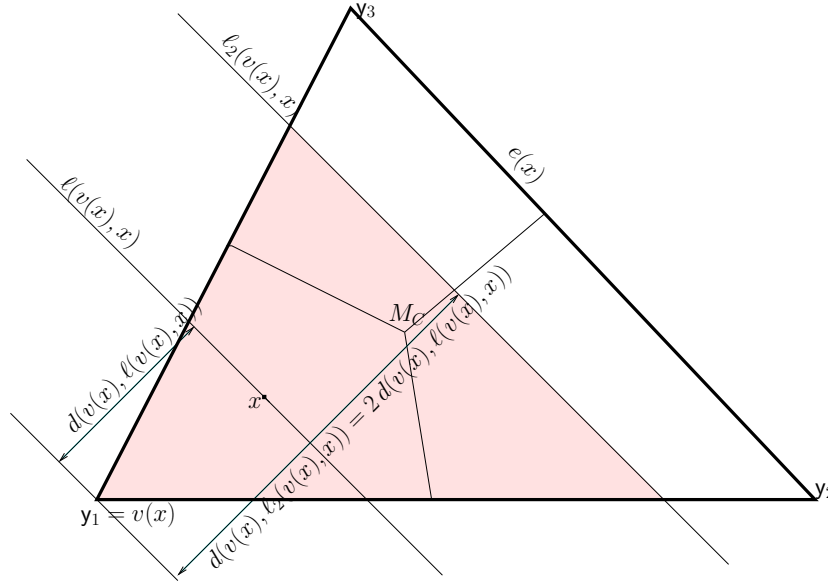


Figure 15: Construction of proximity region, $N_{PE}^2(x)$ (shaded region) for an $x \in R_{CM}(y_1)$.

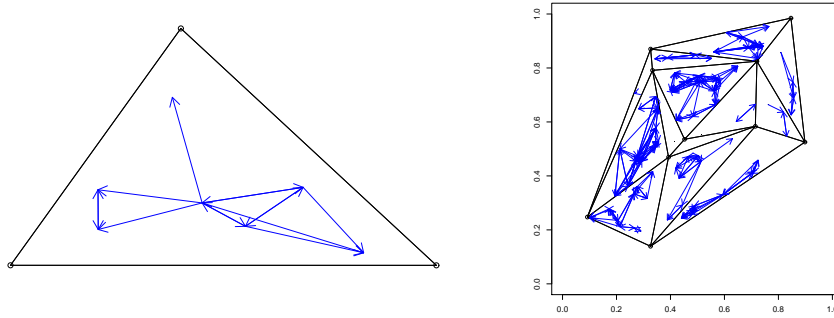


Figure 16: The arcs for $N_{PE}^{r=2}(x, M_C)$ the 7 \mathcal{X} points in Figure 13, and the arcs for $N_{PE}^{r=2}(x, M_C)$ for the 77 \mathcal{X} points that lie in the $\mathcal{C}_H(\mathcal{Y}_{10})$ in Figure 2.

The functional form of $N_{PE}^r(x, M)$ for $x = (x_0, y_0) \in T_b$ is given by

$$N_{PE}^r(x, M) = \left\{ (x, y) \in T_b : y \leq \frac{r(y_0(1-c_1)+c_2x_0)-c_2x}{1-c_1} \right\} \text{ for } x \in R_M(y_1),$$

$$N_{PE}^r(x, M) = \left\{ (x, y) \in T_b : y \leq \frac{r(y_0c_1-c_2(x_0+1))-c_2+c_2x}{c_1} \right\} \text{ for } x \in R_M(y_2),$$

$$N_{PE}^r(x, M) = \{(x, y) \in T_b : y \geq 2y_0 - c_2(r-1)\} \text{ for } x \in R_M(y_3).$$

Of particular interest is $N_{PE}^r(x, M)$ with any M and $r \in \{\sqrt{2}, 3/2, 2\}$. For $r = \sqrt{2}$, $\ell(v(x), x)$ divides $T_{\sqrt{2}}(x)$ into two regions of equal area, hence $N_{PE}^{\sqrt{2}}(x, M)$ is also referred to as *double-area proximity region*. For $r = 2$, $\ell(v(x), x)$ divides the edges of $T_2(x)$ —other than $\ell_r(v(x), x)$ — into two segments of equal length, hence $N_{PE}^2(x, M)$ is also referred to as *double-edge proximity region*. For $r < 3/2$, $\mathcal{R}_S(N_{PE}^r, M_C) = \emptyset$, and for $r > 3/2$, $\mathcal{R}_S(N_{PE}^r, M_C)$ has positive area; for $r = 3/2$, $\mathcal{R}_S(N_{PE}^r, M_C) = \{M_C\}$. Therefore, $r = 3/2$ is the threshold for $N_{PE}^r(\cdot, M_C)$ to satisfy **P6**. Furthermore, $r = 3/2$ is the value at which the asymptotic distribution of the domination number of the PCD based on $N_{PE}^r(\cdot, M_C)$ will be nondegenerate (see Ceyhan (2004) and Ceyhan and Priebe (2005)).

The properties **P1**, **P2**, **P4**, **P5**, and **P7** follow by definition for all M and r . Furthermore **P9** holds, since N_{PE}^r is geometry invariant for uniform data. Property **P5** holds with similarity ratio of $N_{PE}^r(x, M)$ to $T(\mathcal{Y}_3)$: $\frac{\min(d(v(x), e(x)), r d(v(x), \ell(v(x), x)))}{d(v(x), e(x))}$; that is, $N_{PE}^r(x, M)$ is similar to $T(\mathcal{Y}_3)$ with the given ratio.

P6 holds depending on the pair M and r . That is, there exists an $r_o := r_o(M)$ so that $N_{PE}^{r_o}(x, M)$ satisfies **P6** for all $r \leq r_o(M)$, and fails to satisfy otherwise. **P6** fails for all M when $r = \infty$. **P8** holds only when $M = M_C$. With CM -vertex regions, for all $r \in [1, \infty]$, the area $A(N_{PE}^r(x, M_C))$ is a continuous function of $d(\ell_r(v(x), x), v(x))$ which is a continuous function of $d(\ell(v(x), x), v(x))$ which is a continuous function of x .

Moreover, $\Lambda_0(N_{PE}^r, M) = \mathcal{Y}_3$ for all $r \in [1, \infty]$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, since the \mathbb{R}^2 -Lebesgue measure $\lambda(N_{PE}^r(x, M)) = 0$ iff $x \in \mathcal{Y}_3$.

As for **P3**, for $T_2(x) \subseteq T(\mathcal{Y}_3)$ one can loosen the concept of center by treating the line $\ell(v(x), x)$ as the *edge-wise central line*, so **P3** is satisfied in this loose sense for $r = 2$. Notice that x is not the unique center in this sense but a point on a central line. Let M_i , $i \in \{1, 2, 3\}$, be the midpoints of the edges of $T(\mathcal{Y}_3)$, and $T(M_1, M_2, M_3)$ be triangle whose vertices are these midpoints. Then for any $x \in T(M_1, M_2, M_3)$, $N_{PE}^2(x, M) = T(\mathcal{Y}_3)$, so $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S(N_{PE}^2, M)$ where equality holds for $M = M_C$ for all triangles and for $M = M_{CC}$ in non-obtuse triangles (see Figure 17 (left)).

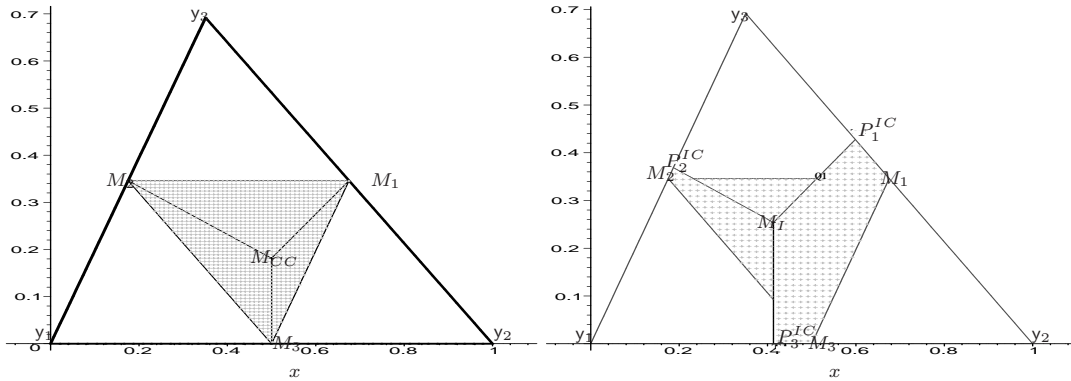


Figure 17: Superset region $\mathcal{R}_S(N_{PE}^2, M_{CC})$ in an acute triangle (left), superset region, $\mathcal{R}_S^\perp(N_{PE}^2, M_I)$ (right)

For an example of double-edge proximity regions $N_{PE}^2(x, M)$ with CC -vertex regions with orthogonal projections, see Figure 18 (top left). Notice that we use the vertex closest to x ; i.e., $\arg\min_{y \in \mathcal{Y}_3} d(x, y)$ for $N_{PE}^2(x, CC)$, i.e. vertex regions $R_{CC}(\cdot)$. Furthermore, if x is close enough to M , it is possible to have

$N_{PE}^2(x, M) = T(\mathcal{Y}_3)$. See Figure 18 (bottom) for an example with CC -vertex regions with orthogonal projections.

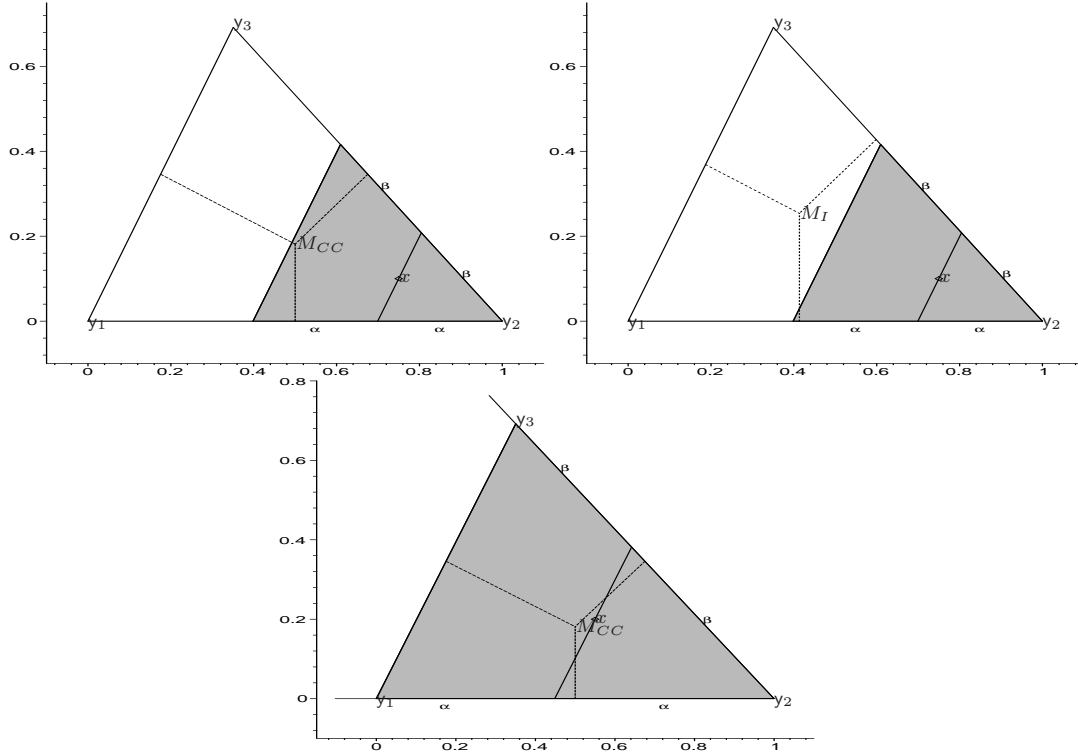


Figure 18: Shaded regions are double-edge proximity regions $N_{PE}^2(x, M) \subseteq T(\mathcal{Y}_3)$ with $M = M_{CC}$ for an $x \in R_{CC}^\perp(y_2)$ (top left), with $M = M_I$ for an $x \in R_{IC}^\perp(y_2)$ (top right). Depicted in the bottom is an example of $N_{PE}^2(x, M) = T(\mathcal{Y}_3)$ with $M = M_{CC}$ for an $x \in R_{CC}^\perp(y_2)$.

In non-obtuse triangles, $\mathcal{R}_S(N_{PE}^2, M_{CC}) = T(M_1, M_2, M_3)$. But, in obtuse triangles, $\mathcal{R}_S(N_{PE}^2, M_{CC}) \supsetneq T(M_1, M_2, M_3)$ and is a quadrilateral. The functional form of the superset region, $\mathcal{R}_S(N_{PE}^2, M)$, in T_b is given by

$$\mathcal{R}_S(N_{PE}^2, M) = \left\{ (x, y) \in R_M(y_1) : y \geq \frac{c_2(1-rx)}{r(1-c_1)} \right\} \cup \left\{ (x, y) \in R_M(y_2) : y \geq \frac{c_2(r(x-1)+1)}{rc_1} \right\} \cup \left\{ (x, y) \in R_M(y_3) : y \leq c_2 \frac{r-1}{r} \right\},$$

and the functional form of $T(M_1, M_2, M_3)$ in T_b is given by

$$T(M_1, M_2, M_3) = \left\{ (x, y) \in T_b : y \leq \frac{c_2}{2}; y \geq \frac{c_2(-1+2x)}{2c_1}; y \geq \frac{c_2(1-2x)}{2(1-c_1)} \right\}.$$

Let $\mathcal{R}_S^\perp(N_{PE}^2, M)$ be the superset region for N_{PE}^2 based on M -vertex regions with orthogonal projections. See Figure 17 for the superset region $\mathcal{R}_S^\perp(N_{PE}^2, M_I)$. Again $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S^\perp(N_{PE}^2, M_I)$ for all $T(\mathcal{Y}_3)$ with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle. For $N_{PE}^2(\cdot, M_C)$ constructed using the median lines $\mathcal{R}_S(N_{PE}^2, M_C) = T(M_1, M_2, M_3)$ and for $N_{PE}^2(\cdot, M_C)$ constructed by the orthogonal projections, $\mathcal{R}_S^\perp(N_{PE}^2, M_C) \supseteq T(M_1, M_2, M_3)$ with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle. An example

of double-edge proximity regions is given in Figure 18 (top right) where IC -vertex regions with orthogonal projections to the edges is used. We could also use IC -vertex regions obtained by inner angle bisectors. Note also that the superset region $\mathcal{R}_S^\perp(N_{PE}^2, M_I)$ is as in Figure 17. Again $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S^\perp(N_{PE}^2, M_I)$ for all $T(\mathcal{Y}_3)$ and $T(M_1, M_2, M_3) = \mathcal{R}_S^\perp(N_{PE}^2, M_I)$ iff $T(\mathcal{Y}_3)$ is an equilateral triangle.

For $r = \sqrt{2}$, one can loosen the concept of center by treating the line $\ell(v(x), x)$ as the *area-wise central line* in $N_{PE}^{\sqrt{2}}(x, M)$, so **P3** is satisfied in this loose sense. For an example of $N_{PE}^{\sqrt{2}}(x, M)$ with CC -vertex regions with orthogonal projections, see Figure 19 (top left). $\mathcal{R}_S(N_{PE}^{\sqrt{2}}, M_{CC})$ has positive area; see Figure 20. An example of double-area proximity region with IC -vertex regions is given at Figure 19 (top right) with orthogonal projections to the edges. Note that if x is close enough to M , it is possible to have $N_{PE}^{\sqrt{2}}(x, M) = T(\mathcal{Y}_3)$. See Figure 19 (bottom) with CC -vertex regions with orthogonal projections. We could also use IC -vertex regions obtained by inner angle bisectors.

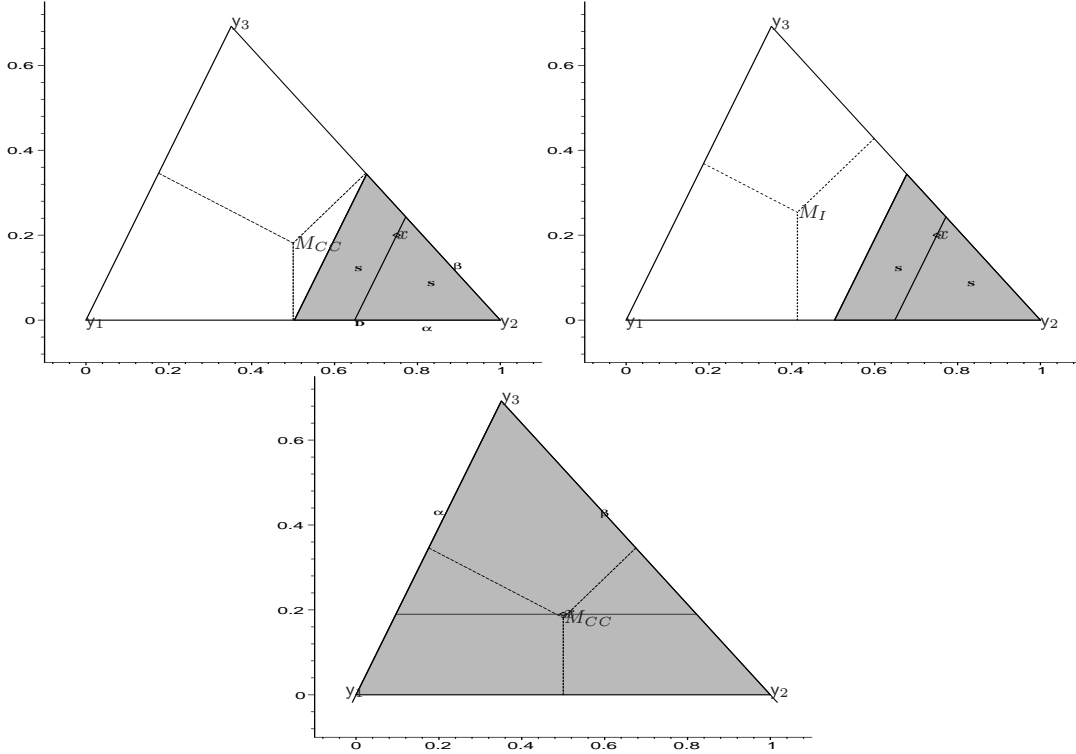


Figure 19: Shaded regions are double-area proximity regions $N_{PE}^{\sqrt{2}}(x, M) \subsetneq T(\mathcal{Y}_3)$ with $M = M_{CC}$ for an $x \in R_{CC}^\perp(y_2)$ (top left), with $M = M_I$ for an $x \in R_{IC}^\perp(y_2)$ (top right) and $N_{PE}^{\sqrt{2}}(x, M) = T(\mathcal{Y}_3)$ for an $x \in R_{CC}^\perp(y_2)$ with $M = M_{CC}$ (bottom).

Note also that $N_{PE}^{\sqrt{2}}(x, M_I) = T(\mathcal{Y}_3)$ might occur if x is close enough to M_I when $M_I \notin \mathcal{T}^r$. Let h_j be the altitude of $T(\mathcal{Y}_3)$ at vertex y_j , for $j = 1, 2, 3$. If $r_{ic} < \left(\frac{\sqrt{2}-1}{\sqrt{2}}\right) h_j$ for some $j \in \{1, 2, 3\}$, then $\mathcal{R}_S(N_{PE}^{\sqrt{2}}, M_I)$ has positive area. See Figure 20 where the superset region is barely noticeable. If $r_{ic} \geq \left(\frac{\sqrt{2}-1}{\sqrt{2}}\right) \max_{j \in \{1, 2, 3\}} \{h_j\}$, then $\mathcal{R}_S(N_{PE}^{\sqrt{2}}, M_I)$ has zero area. In T_b , $r_{ic} \geq \left(\frac{\sqrt{2}-1}{\sqrt{2}}\right) \max\{h_1, h_3\}$ always hold, but $r_{ic} \geq \left(\frac{\sqrt{2}-1}{\sqrt{2}}\right) h_2$ holds iff $c_1^2 + c_2^2 < (\sqrt{2}-1)^2 \left(1 + \sqrt{(1-c_1)^2 + c_2^2}\right)^2$ iff $|e_2| < (\sqrt{2}-1)(1+|e_1|)$.

In $T(\mathcal{Y}_3)$, drawing the lines $q_i(r, x)$ such that $d(y_i, e_i) = r d(q_i(r, x), y_i)$ for $i \in \{1, 2, 3\}$ yields a triangle,

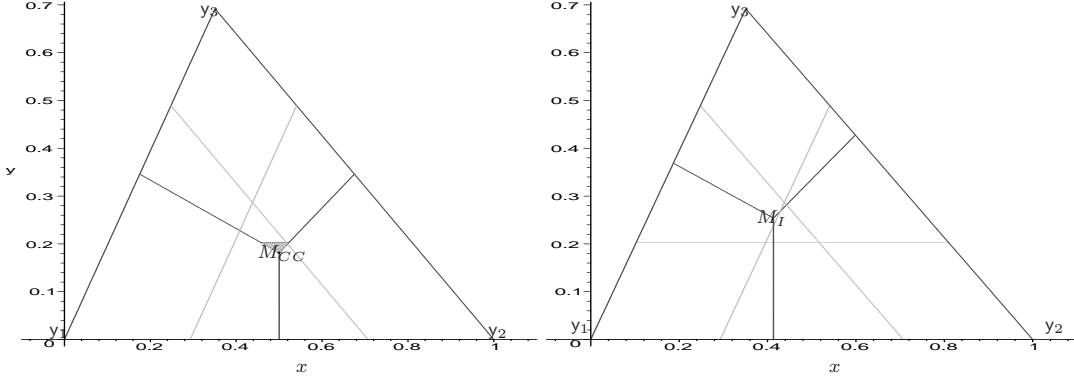


Figure 20: The superset region, $\mathcal{R}_S(N_{PE}^{\sqrt{2}}, M)$ with $M = M_{CC}$ (left) and $M = M_I$ (right).

\mathcal{T}^r , for $r < 3/2$. See Figure 21 for \mathcal{T}^r with $r = \sqrt{2}$. The functional form of \mathcal{T}^r in T_b is

$$\begin{aligned} \mathcal{T}^r &= \left\{ (x, y) \in T_b : y \geq \frac{c_2(r-1)}{r}; y \leq \frac{c_2(1-rx)}{r(1-c_1)}; y \leq \frac{c_2(r(x-1)+1)}{rc_1} \right\} = T(t_1(r), t_2(r), t_3(r)) \\ &= T\left(\left(\frac{(r-1)(1+c_1)}{r}, \frac{c_2(r-1)}{r}\right), \left(\frac{2-r+c_1(r-1)}{r}, \frac{c_2(r-1)}{r}\right), \left(\frac{c_1(2-r)+r-1}{r}, \frac{c_2(r-2)}{r}\right)\right). \end{aligned} \quad (1)$$

There is a crucial difference between \mathcal{T}^r and $T(M_1, M_2, M_3)$: $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S(N_{PE}^r, M)$ for all M and $r \geq 2$, but $(\mathcal{T}^r)^o$ and $\mathcal{R}_S(N_{PE}^r, M)$ are disjoint regions for all M and r . So if $M \in (\mathcal{T}^r)^o$, then $\mathcal{R}_S(N_{PE}^r, M) = \emptyset$; if $M \in \partial(\mathcal{T}^r)$, then $\mathcal{R}_S(N_{PE}^r, M) = \{M\}$; and if $M \notin \mathcal{T}^r$, then $\mathcal{R}_S(N_{PE}^r, M)$ has positive area. Thus $N_{PE}^r(\cdot, M)$ fails to satisfy **P6** if $M \notin \mathcal{T}^r$. The triangle \mathcal{T}^r defined above plays a crucial role in the analysis of the distribution of the domination number of the proportional-edge PCD. In fact, it has been shown that for $M \in \{t_1(r), t_2(r), t_3(r)\}$ there exists a specific value of r for which the asymptotic distribution of the domination number is non-degenerate (Ceyhan and Priebe (2007)). The superset region $\mathcal{R}_S(N_{PE}^r, M)$ will be important for both the domination number and the relative density of the corresponding PCDs.

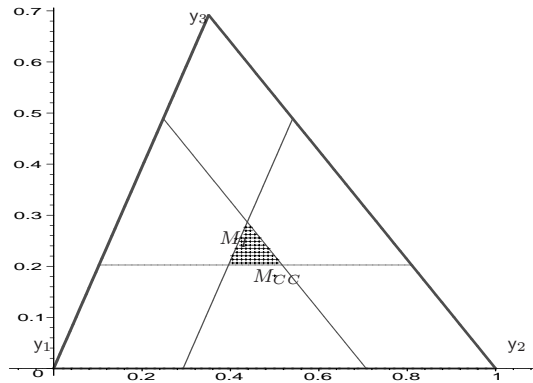


Figure 21: The triangle $\mathcal{T}^{r=\sqrt{2}}$.

In non-acute triangles, the above condition holds for $M = M_{CC}$, since in right and obtuse triangles, $M_{CC} \notin T(\mathcal{Y}_3)^o \Rightarrow M_{CC} \notin \mathcal{T}^r$ (since $T(\mathcal{Y}_3) \supset \mathcal{T}^r$). For an acute basic triangle, if $y_{cc} < \frac{c_2(\sqrt{2}-2x_{cc})}{2(1-c_1)}$

holds, then $M_{CC} \notin \mathcal{T}^{r=\sqrt{2}}$ and the superset region for such triangles is the triangle with vertices M_{CC} , $\left(\frac{-c_2^2 + c_1^2 - 1 + c_2^2\sqrt{2}}{2(-1 + c_1)}, c_2 - \frac{c_2}{\sqrt{2}}\right)$, and $\left(\frac{-c_2^2 + c_1^2 + c_2^2\sqrt{2}}{2c_1}, c_2 - \frac{c_2}{\sqrt{2}}\right)$.

Remark 8.1. In terms of the properties stated in Section 4, $N_{PE}^{3/2}(\cdot, M_C)$ is the most appealing proximity map in the family $\mathcal{N}_{PE}^r := \{N_{PE}^r(\cdot, M) : r \in [1, \infty], M \in \mathbb{R}^2 \setminus \mathcal{Y}_3\}$. It is also noteworthy that the asymptotic distribution of the domination number of the PCD based on $N_{PE}^{3/2}(\cdot, M_C)$ is nondegenerate. \square

Remark 8.2.

- For $r_1 \leq r_2$, $N_{PE}^{r_1}(x, M) \subseteq N_{PE}^{r_2}(x, M)$ for all $x \in T(\mathcal{Y}_3)$. For $r_1 < r_2$, $N_{PE}^{r_1}(x, M) \subseteq N_{PE}^{r_2}(x, M)$ with equality holding for only $x \in \mathcal{Y}_3$ or $x \in \mathcal{R}_S(N_{PE}^{\min(r_1, r_2)}, M)$.
- For $3/2 < r_1 < r_2$, $\mathcal{R}_S(N_{PE}^{r_1}, M) \subsetneq \mathcal{R}_S(N_{PE}^{r_2}, M)$ and $\mathcal{R}_S(N_{PE}^r, M) = \emptyset$ for $r < 3/2$.
- For $r_1 < r_2$, $A(N_{PE}^{r_1}(X, M)) \leq^{ST} A(N_{PE}^{r_2}(X, M))$ for X from a continuous distribution on $T(\mathcal{Y}_3)$ where \leq^{ST} stands for “stochastically smaller than”. \square

8.2.1 Extension of N_{PE}^r to Higher Dimensions

The extension to \mathbb{R}^d for $d > 2$ is straightforward. The extension with $M = M_C$ is given her, but the extension for general M is similar. Let $\mathcal{Y}_{d+1} = \{y_1, y_2, \dots, y_{d+1}\}$ be $d+1$ points that do not lie on the same $d-1$ -dimensional hyperplane. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y}_{d+1})$. A simplex is the simplest polytope in \mathbb{R}^d having $d+1$ vertices, $d(d+1)/2$ edges and $d+1$ faces of dimension $(d-1)$. For $r \in [1, \infty]$, define the proximity map as follows. Given a point x in $\mathfrak{S}(\mathcal{Y}_{d+1})$, let $v := \operatorname{argmin}_{y \in \mathcal{Y}_{d+1}} V(Q_y(x))$ where $Q_y(x)$ is the polytope with vertices being the $d(d+1)/2$ midpoints of the edges, the vertex v and x and $V(\cdot)$ is the d -dimensional volume functional. That is, the vertex region for vertex v is the polytope with vertices given by v and the midpoints of the edges. Let $v(x)$ be the vertex in whose region x falls. If x falls on the boundary of two vertex regions, $v(x)$ is assigned arbitrarily. Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and $\Upsilon(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ which contains x . Let $d(v(x), \Upsilon(v(x), x))$ be the (perpendicular) Euclidean distance from $v(x)$ to $\Upsilon(v(x), x)$. For $r \in [1, \infty]$, let $\Upsilon_r(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ such that

$$d(v(x), \Upsilon_r(v(x), x)) = r d(v(x), \Upsilon(v(x), x))$$

and

$$d(\Upsilon(v(x), x), \Upsilon_r(v(x), x)) < d(v(x), \Upsilon_r(v(x), x)).$$

Let $\mathfrak{S}_r(x)$ be the polytope similar to and with the same orientation as \mathfrak{S} having $v(x)$ as a vertex and $\Upsilon_r(v(x), x)$ as the opposite face. Then the proximity region $N_{PE}^r(x, M_C) := \mathfrak{S}_r(x) \cap \mathfrak{S}(\mathcal{Y}_{d+1})$. Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M_C)$.

8.3 Central Similarity Proximity Maps

The other type of triangular proximity map introduced is the central similarity proximity map. This will turn out to be the most appealing proximity map in terms of the properties in Section 4. Furthermore, the relative density of the corresponding PCD will have mathematical tractability (Ceyhan et al. (2007)). Alas, the distribution of the domination number of the associated PCD is still an open problem (Ceyhan (2004)).

For $\tau \in [0, 1]$, define $N_{CS}^\tau(\cdot, M) := N(\cdot, M; \tau, \mathcal{Y}_3)$ to be the *central similarity proximity map* with M -edge regions as follows; see also Figure 22 with $M = M_C$. For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $e(x)$ be the edge in whose region x falls; i.e., $x \in R_M(e(x))$. If x falls on the boundary of two edge regions, $e(x)$ is assigned to x arbitrarily. For $\tau \in (0, 1]$, the central similarity proximity region $N_{CS}^\tau(x, M)$ is defined to be the triangle $T_\tau(x)$ with the following properties:

- (i) $T_\tau(x)$ has edges $e_i^\tau(x)$ parallel to e_i for $i \in \{1, 2, 3\}$, and for $x \in R_M(e(x))$, $d(x, e^\tau(x)) = \tau d(x, e(x))$ and $d(e^\tau(x), e(x)) \leq d(x, e(x))$ where $d(x, e(x))$ is the Euclidean (perpendicular) distance from x to $e(x)$;
- (ii) $T_\tau(x)$ has the same orientation as and is similar to $T(\mathcal{Y}_3)$;
- (iii) x is the same type of center of $T_\tau(x)$ as M is of $T(\mathcal{Y}_3)$.

Note that (i) implies the parametrization of the PCD, (ii) explains “similarity”, and (iii) explains “central” in the name, *central similarity proximity map*. For $\tau = 0$, $N_{CS}^{\tau=0}(x, M) := \{x\}$ for all $x \in T(\mathcal{Y}_3)$. For $x \in \partial(T(\mathcal{Y}_3))$, $N_{CS}^\tau(x, M) := \{x\}$ for all $\tau \in [0, 1]$.

Notice that by definition $x \in N_{CS}^\tau(x, M)$ for all $x \in T(\mathcal{Y}_3)$. Furthermore, $\tau \leq 1$ implies that $N_{CS}^\tau(x, M) \subseteq T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3)$ and $M \in T(\mathcal{Y}_3)^o$. For all $x \in T(\mathcal{Y}_3)^o \cap R_M(e(x))$, the edges $e^\tau(x)$ and $e(x)$ are coincident iff $\tau = 1$. See Figure 23 for the arcs based on $N_{CS}^{\tau=1}(x, M_C)$ for 20 \mathcal{X} points in the one triangle case.

Notice that $X_i \stackrel{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional pdf f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of $N_{CS}^\tau(\cdot)$ — X falls on the boundary of two edge regions — occurs with probability zero. Note that for such an F , $N_{CS}^\tau(X, M)$ is a triangle for $\tau > 0$ a.s.

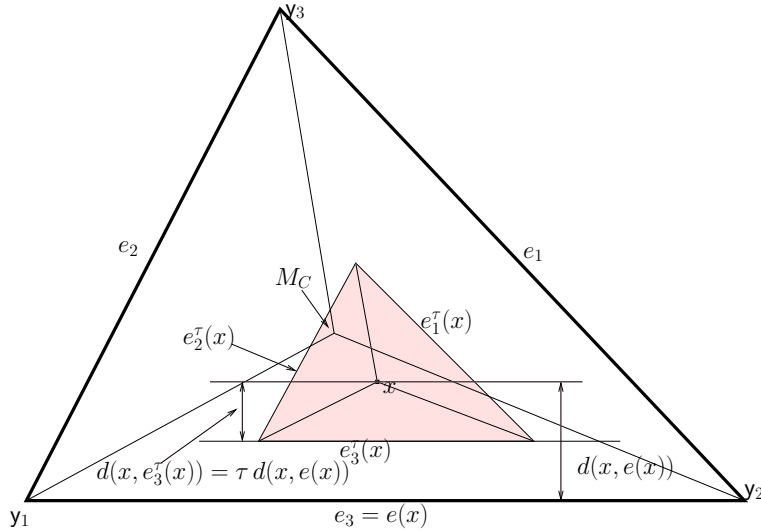


Figure 22: Construction of central similarity proximity region, $N_{CS}^{\tau=1/2}(x, M_C)$ (shaded region).

Notice that central similarity proximity maps are defined with M -edge regions for $M \in T(\mathcal{Y}_3)^o$. Among the four centers considered in Section 6, M_C and M_I are inside the triangle, so they can be used in construction of the central similarity proximity map.

With $M = M_C$, for $x \in R_{CM}(e)$, the similarity ratio of $N_{CS}^\tau(x, M_C)$ to $T(\mathcal{Y}_3)$ is $d(x, e_\tau)/d(M_C, e)$.

See Figure 22 for $N_{CS}^{\tau=1/2}(x, M_C)$ with $e = e_3$ and Figure 24 for $N_{CS}^{\tau=1}(x, M_C)$ with $e = e_3$. The functional form of $N_{CS}^\tau(x, M_C)$ for an $x = (x_0, y_0) \in R_{CM}(e)$ is as follows:

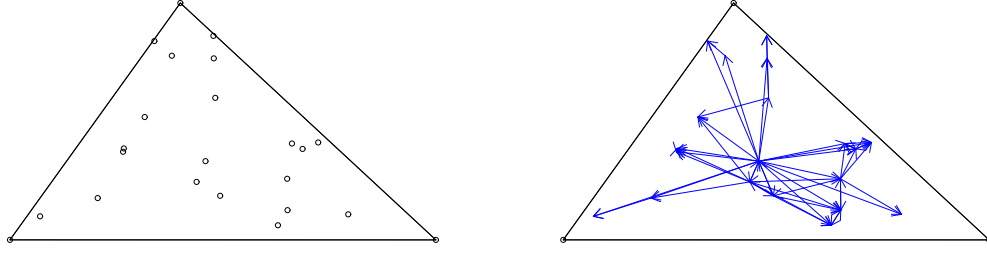


Figure 23: A realization of 20 \mathcal{X} points generated iid $\mathcal{U}(T(\mathcal{Y}_3))$ (left) and the corresponding arcs for $N_{CS}^{\tau=1}(x, M_C)$ (right).

For $x \in R_{CM}(e_1)$,

$$N_{CS}^{\tau}(x, M_C) = \left\{ (x, y) \in T_b : y \geq y_0 + \tau(y_0(1 - c_1)), y \leq \frac{(1 - \tau)(y_0(1 - c_1) - x_0 c_2) + c_2(\tau - x)}{1 - c_1}; \right. \\ \left. y \leq \frac{y_0(c_1(1 + \tau) - \tau) - x_0 c_2(1 + \tau) + c_2(\tau + x)}{c_1} \right\}.$$

For $x \in R_{CM}(e_2)$,

$$N_{CS}^{\tau}(x, M_C) = \left\{ (x, y) \in T_b : y \geq y_0 + \tau(y_0 c_1 - c_2 x_0), y \leq \frac{(1 - \tau)(y_0 c_1 - x_0 c_2) + c_2 x}{c_1}; \right. \\ \left. y \leq \frac{y_0(1 - c_1(1 + \tau)) + x_0 c_2(1 + \tau)}{1 - c_1} \right\}.$$

For $x \in R_{CM}(e_3)$,

$$N_{CS}^{\tau}(x, M_C) = \left\{ (x, y) \in T_b : y \geq y_0(1 - \tau); y \leq \frac{(y_0(\tau + c_1) + c_2(x - x_0))}{c_1}; \right. \\ \left. y \leq \frac{y_0(1 - c_1 + \tau) - c_2(x - x_0)}{1 - c_1} \right\}.$$

8.3.1 IC-Central Similarity Proximity Regions

With $M = M_I$, the similarity ratio is $d(x, e_{\tau})/d(M_I, e)$. See Figure 24 for $N_{CS}^{\tau=1}(x, M_I)$ with $e = e_3$. The functional form of $N_{CS}^{\tau}(x, M_I)$ for an $x = (x_0, y_0) \in R_{IC}(e)$ is as follows:

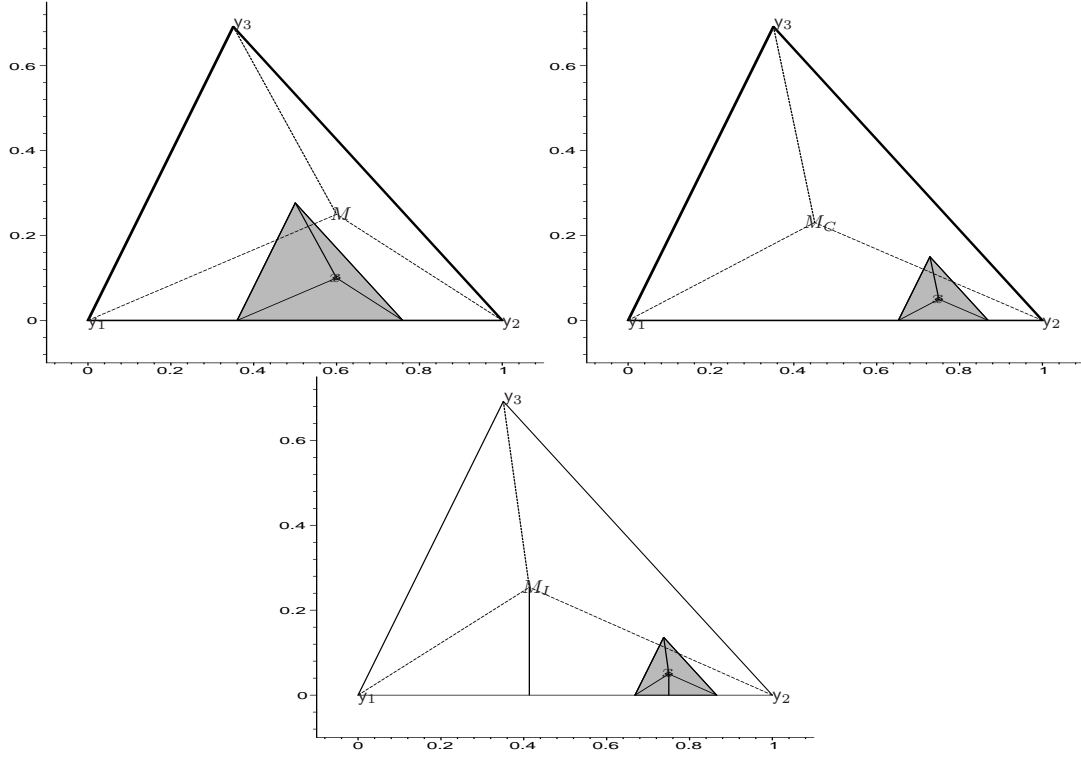


Figure 24: $N_{CS}^{\tau=1}(x, M)$ with an $x \in R_M(e_3)$ (top left); $N_{CS}^{\tau=1}(x, M_C)$ with an $x \in R_{M_C}(e_3)$ (top right); and $N_{CS}^{\tau=1}(x, M_{IC})$ with an $x \in R_{M_{IC}}(e_3)$ (top right).

If $x \in R_{IC}(e_1)$, then

$$N_{CS}^{\tau}(x, M_I) = \left\{ (x, y) \in T_b : y \geq \frac{y_0 \left(\sqrt{c_2^2 + (1 - c_1)^2} + \tau (1 - c_1) \right) + c_2 \tau (x_0 - \tau)}{\sqrt{c_2^2 + (1 - c_1)^2}}; \right. \\ \left. y \leq \frac{(1 - \tau) (y_0 (1 - c_1) + x_0 c_2) + c_2 (\tau - x)}{1 - c_1}; y \leq \frac{\left(\sqrt{c_2^2 + (1 - c_1)^2} + \tau \sqrt{c_1^2 + c_2^2} \right) (y_0 c_1 - x_0 c_2) - y_0 \tau \sqrt{c_1^2 + c_2^2}}{c_1 \sqrt{c_2^2 + (1 - c_1)^2}} \right\}.$$

If $x \in R_{IC}(e_2)$, then

$$N_{CS}^{\tau}(x, M_I) = \left\{ (x, y) \in T_b : y \geq \frac{y_0 \left(\sqrt{c_1^2 + c_2^2} + \tau c_1 \right) - \tau c_2 x_0}{\sqrt{c_1^2 + c_2^2}}; y \leq \frac{(1 - \tau) (c_1 y_0 - c_2 x_0) + c_2 x}{c_1}; \right. \\ \left. y \leq \frac{\left(\sqrt{c_1^2 + c_2^2} + \tau \sqrt{c_2^2 + (1 - c_1)^2} \right) (x_0 c_2 - y_0 c_1) + \sqrt{c_1^2 + c_2^2} (y_0 - c_2 x)}{(1 - c_1) \sqrt{c_1^2 + c_2^2}} \right\}.$$

If $x \in R_{IC}(e_3)$, then

$$N_{CS}^\tau(x, M_I) = \left\{ (x, y) \in T_b : y \geq y_0(1 - \tau); y \leq \frac{y_0(c_1 + \tau\sqrt{c_1^2 + c_2^2}) + c_2(x - x_0)}{c_1}, \right. \\ \left. y \leq \frac{y_0(1 - c_1 + \tau\sqrt{c_2^2 + (1 - c_1)^2}) + c_2(x_0 - x)}{1 - c_1} \right\}.$$

$N_{CS}^\tau(\cdot, M_I)$ also satisfies all the properties **P1-P9**.

Remark 8.3. For acute triangles we could use CC or OC -edge regions in central similarity proximity regions which will also satisfy properties **P1-P9**. But for obtuse triangles, **P2** is not satisfied and edge regions are not defined in a natural manner.

In general for M -central similarity proximity regions, the similarity ratio is $d(x, e_\tau(x))/d(M, e(x))$. See Figure 24 for $N_{CS}^{\tau=1}(x, M)$ with $e = e_3$. The functional form of $N_{CS}^\tau(x, M)$ for an $x = (x_0, y_0) \in R_M(e)$ is as follows:

If $x \in R_M(e_1)$, then

$$N_{CS}^\tau(x, M) = \left\{ (x, y) \in T_b : y \geq \frac{y_0(m_2(1 - \tau)(c_1 - 1) + c_2(1 - m_1)) - \tau m_2 c_2(1 - x_0)}{m_2(1 - c_1) + c_2(1 - m_1)}; \right. \\ \left. y \leq \frac{(1 - \tau)(y_0(1 - c_1) + x_0 c_2) + c_2(\tau - x)}{1 - c_1}; y \leq \left[y_0(c_1 m_2(1 - \tau)(1 - c_1) + m_1 c_2(\tau + c_1) - c_1 c_2(1 + \tau m_1)) + x_0(c_2^2(1 - m_1(1 - \tau)) + c_2 m_2(c_1(1 - \tau) - c_2)) + \tau c_2(m_2 c_1 - m_1 c_2) + (c_2 m_2(1 - c_1) - c_2^2(1 - m_1))x \right] / \left[c_1(m_2(1 - c_1) - c_1(1 - m_1)) \right] \right\}.$$

If $x \in R_M(e_2)$, then

$$N_{CS}^\tau(x, M) = \left\{ (x, y) \in T_b : y \geq \frac{y_0(c_2 m_1 - c_1 m_2) - \tau m_2(c_2 x_0 - c_1 y_0)}{m_1 c_2 - c_1 m_2}; y \leq \frac{(1 - \tau)(y_0 c_1 - x_0 c_2) + c_2 x}{c_1}; \right. \\ \left. y \leq \left[y_0(c_1 m_2 - c_2 m_1) + \tau c_1(c_2(1 - m_1) - m_2(1 - c_1)) + x_0(c_2^2(-\tau - m_1(1 - \tau)) + m_2 c_2(\tau(1 - c_1) + c_1)) + c_2(m_1 c_2 - m_2 c_1)x \right] / \left[(1 - c_1)(c_1 m_2 - c_2 m_1) \right] \right\}.$$

If $x \in R_M(e_3)$, then

$$N_{CS}^\tau(x, M) = \left\{ (x, y) \in T_b : y \geq y_0(1 - \tau); y \leq \frac{y_0(c_1 m_2(1 - \tau) + c_2 \tau m_1) + c_2 m_2(x - x_0)}{c_1 m_2}; \right. \\ \left. y \leq \frac{y_0(m_2(1 - c_1)(1 - \tau) + c_2 \tau) + c_2 m_2(x_0 - c_2)}{(1 - c_1)m_2} \right\}.$$

Notice that $N_{CS}^\tau(\cdot, M)$ also satisfies properties **P1-P9**. M -central similarity proximity regions with $m_1 < c_1$ can be defined in a similar fashion. Furthermore, $\Lambda_0(N_{CS}^\tau(\cdot, M)) = \partial(T(\mathcal{Y}_3))$ for all $\tau \in (0, 1]$ and $\Lambda_0(N_{CS}^{\tau=0}(\cdot, M)) = T(\mathcal{Y}_3)$, since $\lambda(N_{CS}^\tau(x)) = 0$ iff $x \in e_i$ for $i \in \{1, 2, 3\}$ or $\tau = 0$.

Remark 8.4. Among the family $\mathcal{N}_{CS}^\tau := \{N_{CS}^\tau(\cdot, M) : \tau \in [0, 1], M \in T(\mathcal{Y}_3)^o\}$, every $N_{CS}^\tau(\cdot, M)$ with $\tau \in (0, 1]$ satisfies all the properties in Section 4. \square

Remark 8.5.

- For $\tau_1 \leq \tau_2$, $N_{CS}^{\tau_1}(x, M) \subseteq N_{CS}^{\tau_2}(x, M)$ for all $x \in T(\mathcal{Y}_3)$. For $\tau_1 < \tau_2$, $N_{CS}^{\tau_1}(x, M) \subseteq N_{CS}^{\tau_2}(x, M)$ with equality holding only for $x \in \partial(T(\mathcal{Y}_3))$.

- The superset region $\mathcal{R}_S(N_{CS}^\tau, M) = \emptyset$ for $\tau \in [0, 1)$ and $\mathcal{R}_S(N_{CS}^{\tau=1}, M) = \{M\}$.
- For $\tau_1 < \tau_2$, $A(N_{CS}^{\tau_1}(X, M)) \leq^{ST} A(N_{CS}^{\tau_2}(X, M))$ for X from a continuous distribution on $T(\mathcal{Y}_3)$. \square

8.3.2 Extension of N_{CS}^τ to Higher Dimensions

The extension of N_{CS}^τ to \mathbb{R}^d for $d > 2$ is straightforward. the extension for $M = M_C$ is described, the extension for general M is similar. Let $\mathcal{Y}_{d+1} = \{y_1, y_2, \dots, y_{d+1}\}$ be $d + 1$ points that do not lie on the same $(d - 1)$ -dimensional hyperplane. Denote the simplex formed by these $d + 1$ points as $\mathfrak{S}(\mathcal{Y}_{d+1})$. For $\tau \in (0, 1]$, define the central similarity proximity map as follows. Let φ_i be the face opposite vertex y_i for $i \in \{1, 2, \dots, (d + 1)\}$, and “face regions” $R_{CM}(\varphi_1), \dots, R_{CM}(\varphi_{d+1})$ partition $\mathfrak{S}(\mathcal{Y}_{d+1})$ into $d + 1$ regions, namely the $d + 1$ polytopes with vertices being the center of mass together with d vertices chosen from $d + 1$ vertices. For $x \in \mathfrak{S}(\mathcal{Y}_{d+1}) \setminus \mathcal{Y}_{d+1}$, let $\varphi(x)$ be the face in whose region x falls; $x \in R(\varphi(x))$. If x falls on the boundary of two face regions, $\varphi(x)$ is assigned arbitrarily. For $\tau \in (0, 1]$, the central similarity proximity region $N_{CS}^\tau(x, M_C) = \mathfrak{S}_\tau(x)$ is defined to be the simplex $\mathfrak{S}_\tau(x)$ with the following properties:

- (i) $\mathfrak{S}_\tau(x)$ has faces $\varphi_i^\tau(x)$ parallel to $\varphi_i(x)$ for $i \in \{1, 2, \dots, (d+1)\}$, and for $x \in R_{CM}(\varphi(x))$, $\tau d(x, \varphi(x)) = d(\varphi^\tau(x), x)$ where $d(x, \varphi(x))$ is the Euclidean (perpendicular) distance from x to $\varphi(x)$;
- (ii) $\mathfrak{S}_\tau(x)$ has the same orientation as and similar to $\mathfrak{S}(\mathcal{Y}_{d+1})$;
- (iii) x is the center of mass of $\mathfrak{S}_\tau(x)$, as M_C is of $\mathfrak{S}(\mathcal{Y}_{d+1})$. Note that $\tau > 1$ implies that $x \in N_{CS}^\tau(x)$.

8.4 The Behavior of Proximity Regions

In this section, we provide the conditions for x , which, if satisfied, will imply some sort of increase in the size of the proximity regions we have defined. Let $N(\cdot)$ be any proximity map defined on the measurable space Ω with measure μ , and let $\{x_n\}_{n=1}^\infty$ be a sequence of points in Ω . We say $N(x_n)$ gets larger if $N(x_n) \subseteq N(x_m)$ for $m \geq n$, and $N(x_n)$ gets strictly larger if $N(x_n) \subsetneq N(x_m)$ for $m > n$.

In the following theorems we will assume $\Omega = \mathbb{R}^2$ with μ being the \mathbb{R}^2 -Lebesgue measure λ and M -vertex regions are defined with points $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$.

Theorem 8.6. *For arc-slice proximity regions with M -vertex regions for an $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, as $d(x, y)$ (strictly) increases for x lying on a ray from y in $R_M(y) \setminus \mathcal{R}_S(N_{AS}, M)$, $N_{AS}(x, M)$ gets (strictly) larger.*

Proof: For x, y lying on a ray from y in $R_M(y) \setminus \mathcal{R}_S(N_{AS}, M)$, if $d(x, y) \leq d(y, y)$, then $B(x, r(x)) \subseteq B(y, r(y))$, which implies $N_{AS}(x, M) \subseteq N_{AS}(y, M)$, hence $N_{AS}(x, M)$ gets larger as $d(x, y)$ increases for x lying on a ray from y in $R_M(y) \setminus \mathcal{R}_S(N_{AS}, M)$. The strict version follows similarly. If $x, y \in R_M(y) \cap \mathcal{R}_S(N_{AS}, M)$, then $N_{AS}(x, M) = N_{AS}(y, M) = T(\mathcal{Y}_3)$. \blacksquare

Let $\ell(y, x)$ be the line at x parallel to $e(x)$ for $x \in R_M(y)$ where $e(x)$ is the edge opposite vertex y .

Theorem 8.7. *For the proportional-edge proximity maps with M -vertex regions for an $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, as $d(\ell(y, x), y)$ (strictly) increases for $x \in R_M(y) \setminus \mathcal{R}_S(N_{PE}^r, M)$, $N_{PE}^r(x, M)$ gets (strictly) larger for $r < \infty$.*

Proof: For $x, y \in R_M(y) \setminus \mathcal{R}_S(N_{PE}^r, M)$, if $d(\ell(y, x), y) \leq d(\ell(y, y), y)$, then by definition $N_{PE}^r(x, M) \subseteq N_{PE}^r(y, M)$, hence the result follows. The strict version follows similarly. If $x, y \in R_M(y) \cap \mathcal{R}_S(N_{PE}^r, M)$, then $N_{PE}^r(x, M) = N_{PE}^r(y, M) = T(\mathcal{Y}_3)$, and if $r = \infty$ and $x, y \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, $N_{PE}^r(x, M) = N_{PE}^r(y, M) = T(\mathcal{Y}_3)$. \blacksquare

Note that as $d(\ell(y, x), y)$ increases for $x \in R_M(y)$, $d(\ell(y, x), M)$ decreases, provided that $M \in T(\mathcal{Y}_3)^\circ$ and M -vertex regions are convex.

We define the M -edge regions, $R_M(e)$, with points $M \in T(\mathcal{Y}_3)^\circ$.

Theorem 8.8. For central similarity proximity regions with M -edge regions for an $M \in T(\mathcal{Y}_3)^o$, as $d(x, e)$ (strictly) increases for $x \in R_M(e)$, the area $A(N_{CS}^\tau(x, M))$ (strictly) increases for $\tau \in (0, 1]$.

Proof: For $x, y \in R_M(e)$ and $\tau \in (0, 1]$, if $d(x, e) \leq d(y, e)$ then the similarity ratio of $N_{CS}^\tau(y, M)$ to $T(\mathcal{Y}_3)$ is larger than or equal to that of $N_{CS}^\tau(x, M)$, which in turn implies that $A(N_{CS}^\tau(x, M)) \leq A(N_{CS}^\tau(y, M))$. The strict version follows similarly. ■

Observe that the statement of Theorem 8.8 is about the area $A(N_{CS}^\tau(x, M))$. We need further conditions for $N_{CS}^\tau(x, M)$ to get larger.

Theorem 8.9. Let $\ell_M(y)$ be the line joining M and vertex $y \in \mathcal{Y}_3$. As $d(x, \ell_M(y_j))$ and $d(x, \ell_M(y_k))$ both (strictly) decrease for $x \in R_M(e_l)$ where j, k, l are distinct, $N_{CS}^\tau(x, M)$ (strictly) increases for $\tau \in (0, 1]$.

Proof: Suppose, without loss of generality, that $x, y \in R_M(e_3)$. Consider the set

$$S(e_3, x) := \{y \in R_M(e_3) : d(y, \ell_M(y_1)) \leq d(x, \ell_M(y_1)) \text{ and } d(y, \ell_M(y_2)) \leq d(x, \ell_M(y_2))\},$$

which is a parallelogram. See Figure 25 for an example of $S(e_3, x)$ with $M = M_C$ and $e = e_3$. Given $x, y \in S(e_3, x)$, by construction, $N_{CS}^\tau(x, M) \subseteq N_{CS}^\tau(y, M)$. Then the desired result follows for $\tau \in (0, 1]$. Observe that if x_{n+1} is in $S(e_3, x_n)$, then $d(x_n, \ell_M(y_1))$ and $d(x_n, \ell_M(y_2))$ both decrease. The strict version follows similarly. ■

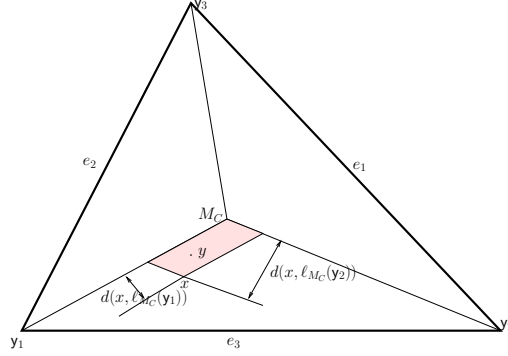


Figure 25: The figure for $x, y \in R_M(e_3)$ described in Theorem 8.9.

Remark 8.10. For $\mathcal{R}_S(N_Y)$ with positive area, by definition, as $x \rightarrow y \in \mathcal{R}_S(N_Y)$, $N_Y(x) \rightarrow T(\mathcal{Y}_3)$ and hence $\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_Y(x)) \in \mathcal{R}_S(N_Y)$ with $\sup_{x \in T(\mathcal{Y}_3)} A(N_Y(x)) = A(T(\mathcal{Y}_3))$.

- As $x \rightarrow M_{CC}$ in a non-obtuse triangle $T(\mathcal{Y}_3)$, $N_{AS}(x, M_{CC}) \rightarrow T(\mathcal{Y}_3)$ and

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{AS}(x, M_{CC})) = M_{CC} \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{AS}(x, M_{CC})) = A(T(\mathcal{Y}_3)).$$

- For $r > 3/2$, as $x \rightarrow y \in (\mathcal{R}_S(N_{PE}^r, M))^o$, $N_{PE}^r(x, M) \rightarrow T(\mathcal{Y}_3)$ hence

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) \in \mathcal{R}_S(N_{PE}^r, M) \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = A(T(\mathcal{Y}_3)).$$

- For $r < 3/2$, if $M \notin \mathcal{T}^r$, then as $x \rightarrow M$, $N_{PE}^r(x, M) \rightarrow T(\mathcal{Y}_3)$ and

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = M \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = A(T(\mathcal{Y}_3)).$$

If $M \in (\mathcal{T}^r)^o$, then as $x \rightarrow M$, $N_{PE}^r(x, M) \rightarrow N_{PE}^r(M, M) \subsetneq T(\mathcal{Y}_3)$, but still

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = M \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = A(N_{PE}^r(M, M)).$$

If $M \in \partial(\mathcal{T}^r)$, then as $x \rightarrow M$, $N_{PE}^r(x, M) \rightarrow N_{PE}^r(M, M) \subseteq T(\mathcal{Y}_3)$, and

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = M \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{PE}^r(x, M)) = A(N_{PE}^r(M, M))$$

which might be $T(\mathcal{Y}_3)$ or a proper subset of $T(\mathcal{Y}_3)$.

- As $x \rightarrow M_C$, $N_{PE}^{3/2}(x, M_C) \rightarrow T(\mathcal{Y}_3)$ and

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{PE}^{3/2}(x, M_C)) = M_C \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{PE}^{3/2}(x, M_C)) = A(T(\mathcal{Y}_3)).$$

- As $x \rightarrow M$, $N_{CS}^{\tau=1}(x, M) \rightarrow T(\mathcal{Y}_3)$ and

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{CS}^{\tau=1}(x, M)) = M \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{CS}^{\tau=1}(x, M)) = A(T(\mathcal{Y}_3)).$$

For $\tau < 1$, as $x \rightarrow M$, $N_{CS}^\tau(x, M) \rightarrow N_{CS}^\tau(M, M)$ and

$$\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{CS}^\tau(x, M)) = M \text{ with } \sup_{x \in T(\mathcal{Y}_3)} A(N_{CS}^\tau(x, M)) = A(N_{CS}^\tau(M, M)). \quad \square$$

Although the comments in the above remark follow by definition, they will be indicative of whether the asymptotic distribution of the domination number of the associated PCD is degenerate or not.

9 Relative Arc Density and Domination Number of PCDs

9.1 Relative Arc Density

The *relative arc density* of a digraph $D = (\mathcal{V}, \mathcal{A})$ of order $|\mathcal{V}| = n$, denoted as $\rho(D)$, is defined as

$$\rho(D) = \frac{|\mathcal{A}|}{n(n-1)}$$

where $|\cdot|$ denotes the set cardinality functional (Janson et al. (2000)). Thus $\rho(D)$ represents the ratio of the number of arcs in the digraph D to the number of arcs in the complete symmetric digraph of order n , which is $n(n-1)$. For brevity of notation we use *relative density* rather than relative arc density henceforth.

If $X_1, \dots, X_n \stackrel{iid}{\sim} F$ the relative density of the associated data-random PCD D , denoted as $\rho(\mathcal{X}_n; h, N_{\mathcal{Y}})$, is a U -statistic,

$$\rho(\mathcal{X}_n; h, N_{\mathcal{Y}}) = \frac{1}{n(n-1)} \sum_{i < j} h(X_i, X_j; N_{\mathcal{Y}}) \quad (2)$$

where

$$\begin{aligned} h(X_i, X_j; N_{\mathcal{Y}}) &= \mathbf{I}\{(X_i, X_j) \in \mathcal{A}\} + \mathbf{I}\{(X_j, X_i) \in \mathcal{A}\} \\ &= \mathbf{I}\{X_j \in N_{\mathcal{Y}}(X_i)\} + \mathbf{I}\{X_i \in N_{\mathcal{Y}}(X_j)\}, \end{aligned} \quad (3)$$

where $\mathbf{I}(\cdot)$ is the indicator function. We denote $h(X_i, X_j; N_Y)$ as h_{ij} for brevity of notation. Since the digraph is asymmetric, h_{ij} is defined as the number of arcs in D between vertices X_i and X_j , in order to produce a symmetric kernel with finite variance (Lehmann (1988)).

The random variable $\rho_n := \rho(\mathcal{X}_n; h, N_Y)$ depends on n and N_Y explicitly and on F implicitly. The expectation $\mathbf{E}[\rho_n]$, however, is independent of n and depends on only F and N_Y :

$$0 \leq \mathbf{E}[\rho_n] = \frac{1}{2} \mathbf{E}[h_{12}] \leq 1 \text{ for all } n \geq 2. \quad (4)$$

The variance $\mathbf{Var}[\rho_n]$ simplifies to

$$0 \leq \mathbf{Var}[\rho_n] = \frac{1}{2n(n-1)} \mathbf{Var}[h_{12}] + \frac{n-2}{n(n-1)} \mathbf{Cov}[h_{12}, h_{13}] \leq 1/4. \quad (5)$$

A central limit theorem for U -statistics (Lehmann (1988)) yields

$$\sqrt{n}(\rho_n - \mathbf{E}[\rho_n]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{Cov}[h_{12}, h_{13}]) \quad (6)$$

provided $\mathbf{Cov}[h_{12}, h_{13}] > 0$. The asymptotic variance of ρ_n , $\mathbf{Cov}[h_{12}, h_{13}]$, depends on only F and N_Y . Thus, we need determine only $\mathbf{E}[h_{12}]$ and $\mathbf{Cov}[h_{12}, h_{13}]$ in order to obtain the normal approximation

$$\rho_n \overset{\text{approx}}{\sim} \mathcal{N}(\mathbf{E}[\rho_n], \mathbf{Var}[\rho_n]) = \mathcal{N}\left(\frac{\mathbf{E}[h_{12}]}{2}, \frac{\mathbf{Cov}[h_{12}, h_{13}]}{n}\right) \text{ for large } n. \quad (7)$$

9.2 Domination Number

In a digraph $D = (\mathcal{V}, \mathcal{A})$, a vertex $v \in \mathcal{V}$ *dominates* itself and all vertices of the form $\{u : vu \in \mathcal{A}\}$. A *dominating set* S_D for the digraph D is a subset of \mathcal{V} such that each vertex $v \in \mathcal{V}$ is dominated by a vertex in S_D . A *minimum dominating set* S_D^* is a dominating set of minimum cardinality and the *domination number* $\gamma(D)$ is defined as $\gamma(D) := |S_D^*|$ (see, Lee (1998)) where $|\cdot|$ denotes the set cardinality functional. See Chartrand and Lesniak (1996) and West (2001) for more on graphs and digraphs. If a minimum dominating set is of size one, we call it a *dominating point*.

Note that for $|\mathcal{V}| = n > 0$, $1 \leq \gamma(D) \leq n$, since \mathcal{V} itself is always a dominating set.

9.3 Asymptotic Distribution of Relative Arc Density of PCDs

By detailed geometric probability calculations, provided in Ceyhan et al. (2006) and Ceyhan et al. (2007) the mean and the asymptotic variance of the relative density of the proportional-edge and central similarity PCDs can explicitly be computed. The central limit theorem for U -statistics then establishes the asymptotic normality under the uniform null hypothesis. These results are summarized in the following theorems.

Theorem 9.1. For $r \in [1, \infty)$,

$$\frac{\sqrt{n}(\rho_n(r) - \mu(r))}{\sqrt{\nu(r)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (8)$$

where

$$\mu(r) = \begin{cases} \frac{37}{216}r^2 & \text{for } r \in [1, 3/2), \\ -\frac{1}{8}r^2 + 4 - 8r^{-1} + \frac{9}{2}r^{-2} & \text{for } r \in [3/2, 2), \\ 1 - \frac{3}{2}r^{-2} & \text{for } r \in [2, \infty), \end{cases} \quad (9)$$

and

$$\nu(r) = \nu_1(r) \mathbf{I}(r \in [1, 4/3)) + \nu_2(r) \mathbf{I}(r \in [4/3, 3/2)) + \nu_3(r) \mathbf{I}(r \in [3/2, 2)) + \nu_4(r) \mathbf{I}(r \in [2, \infty)) \quad (10)$$

with

$$\begin{aligned} \nu_1(r) &= \frac{3007 r^{10} - 13824 r^9 + 898 r^8 + 77760 r^7 - 117953 r^6 + 48888 r^5 - 24246 r^4 + 60480 r^3 - 38880 r^2 + 3888}{58320 r^4}, \\ \nu_2(r) &= \frac{5467 r^{10} - 37800 r^9 + 61912 r^8 + 46588 r^6 - 191520 r^5 + 13608 r^4 + 241920 r^3 - 155520 r^2 + 15552}{233280 r^4}, \\ \nu_3(r) &= -[7 r^{12} - 72 r^{11} + 312 r^{10} - 5332 r^8 + 15072 r^7 + 13704 r^6 - 139264 r^5 + 273600 r^4 - 242176 r^3 \\ &\quad + 103232 r^2 - 27648 r + 8640]/[960 r^6], \\ \nu_4(r) &= \frac{15 r^4 - 11 r^2 - 48 r + 25}{15 r^6}. \end{aligned}$$

For $r = \infty$, $\rho_n(r)$ is degenerate.

Theorem 9.2. For $\tau \in (0, 1]$, the relative density of the central similarity proximity digraph converges in law to the normal distribution; i.e., as $n \rightarrow \infty$,

$$\frac{\sqrt{n}(\rho_n(\tau) - \mu(\tau))}{\sqrt{\nu(\tau)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (11)$$

where

$$\mu(\tau) = \tau^2/6 \quad (12)$$

and

$$\nu(\tau) = \frac{\tau^4(6\tau^5 - 3\tau^4 - 25\tau^3 + \tau^2 + 49\tau + 14)}{45(\tau + 1)(2\tau + 1)(\tau + 2)}. \quad (13)$$

For $\tau = 0$, $\rho_n(\tau)$ is degenerate for all $n > 1$.

9.4 Asymptotic Distribution of Domination Number of PCDs

Recall the triangle \mathcal{T}^r defined in Equation (1) (see also Figure 21 for \mathcal{T}^r with $r = \sqrt{2}$). Let $\gamma_n(r, M) := \gamma(\mathcal{X}_n, N_{PE}^r, M)$ be the domination number of the PCD based on N_{PE}^r with \mathcal{X}_n , a set of iid random variables from $\mathcal{U}(T(\mathcal{Y}_3))$, with M -vertex regions.

The domination number $\gamma_n(r, M)$ of the PCD has the following asymptotic distribution (Ceyhan and Priebe (2007)). As $n \rightarrow \infty$,

$$\gamma_n(r, M) \xrightarrow{\mathcal{L}} \begin{cases} 2 + \text{BER}(1 - p_r) & \text{for } r \in [1, 3/2) \text{ and } M \in \{t_1(r), t_2(r), t_3(r)\}, \\ 1 & \text{for } r > 3/2 \text{ and } M \in T(\mathcal{Y}_3)^o, \\ 3 & \text{for } r \in [1, 3/2) \text{ and } M \in \mathcal{T}^r \setminus \{t_1(r), t_2(r), t_3(r)\}, \end{cases} \quad (14)$$

where $\xrightarrow{\mathcal{L}}$ stands for “convergence in law” and $\text{BER}(p)$ stands for Bernoulli distribution with probability of success p , \mathcal{T}^r and $t_i(r)$ are defined in Equation (1), and for $r \in [1, 3/2)$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$,

$$p_r = \int_0^\infty \int_0^\infty \frac{64 r^2}{9(r-1)^2} w_1 w_3 \exp\left(\frac{4r}{3(r-1)}(w_1^2 + w_3^2 + 2r(r-1)w_1 w_3)\right) dw_3 dw_1, \quad (15)$$

and for $r = 3/2$ and $M = M_C = \{(1/2, \sqrt{3}/6)\}$, $p_r \approx 0.7413$, which is not computed as in Equation (15); for its computation, see Ceyhan and Priebe (2005). For example, for $r = 5/4$ and $M \in \{t_1(r) = (3/10, \sqrt{3}/10), t_2(r) = (7/10, \sqrt{3}/10)\}$,

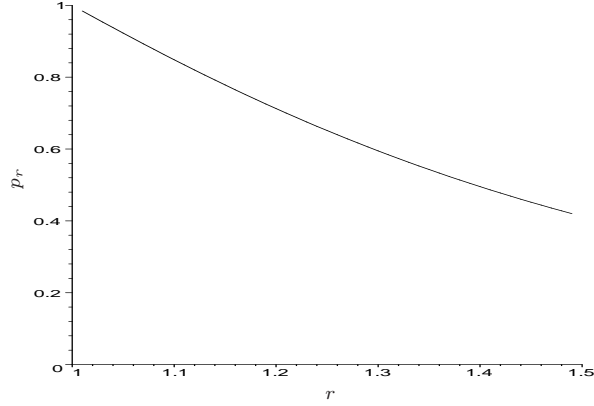


Figure 26: Plotted is the probability $p_r = \lim_{n \rightarrow \infty} P(\gamma_n(r, M) = 2)$ given in Equation (15) as a function of r for $r \in [1, 3/2)$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$.

$p_r \approx 0.6514$. See Figure 26 for the plot of the numerically computed values (i.e., the values computed by numerical integration of the expression in Equation (15)) of p_r as a function of r . Notice that in the nondegenerate case in (14), $\mathbf{E}[\gamma_n(r, M)] = 3 - p_r$ and $\mathbf{Var}[\gamma_n(r, M)] = p_r(1 - p_r)$.

The results in Theorem 5.2 and Corollaries 5.5 and 5.6 also hold for relative arc density and the domination number of PCDs based $N_{\mathcal{Y}}$. That is, we have the following corollary.

Corollary 9.3. *Given any triangle T_o and \mathcal{X}_n a random sample from $\mathcal{U}(T_o)$. Suppose the PCD, D_o is defined in such a way that the ratio of the area of $N(x)$ to the area of the triangle T_o is preserved under the uniformity preserving transformation, then the distributions of the relative arc density and the domination number of D_o are geometry invariant.*

10 Two New Proximity Maps

In this section, we introduce two new proximity maps and investigate their properties.

10.1 Directional-Doubling Proximity Maps

Without loss of generality, we can assume that $T(\mathcal{Y}_3) = T_b$. Partition the triangle T_b by M -edge regions to obtain $R_M(e_i)$ for $i = 1, 2, 3$. For $z \in R_M(e_i)$, directional-doubling proximity map is defined as

$$N_{DD}(z, M) := \{\mathbf{x} \in T_b : d(x, e_i) \leq 2d(z, e_i)\}.$$

See Figure 27 (left) for $M = M_C$. If $z \in e_i$ then $N_{DD}(z, M) := e_i$. Notice that if $z \notin e_i$, then $N_{DD}(z, M)$ is a quadrilateral. Among the properties, **P1** and **P2** follows trivially. The line at $z \in R_M(e_i)$ parallel to e_i divides the region into two pieces (half-way in the perpendicular direction to e_i) so **P3** holds in this special sense. **P4** and **P5** both fail, since $N_{DD}(z, M)$ is a quadrilateral. **P6** holds if $M \in T(M_1, M_2, M_3)$, otherwise it fails since $\mathcal{R}_S(N_{DD}, M)$ will have positive area. **P7** also follows by definition. However, **P8** holds only when $M = M_C$.

Property **P9** follows for N_{DD} , since $N_{DD}(z, M)$ is constructed with the boundary of $T(\mathcal{Y}_3)$ and parallel lines to the edges, by Corollary 5.5, geometry invariance for uniform data follows. That is, the distributions of relative arc density and the domination number of the corresponding PCD do not depend on the geometry of the triangle $T(\mathcal{Y}_3)$. Hence, it suffices to compute them for the standard equilateral triangle only. Furthermore, $\Lambda_0(N_{DD}) = \partial(T(\mathcal{Y}_3))$ since $N_{DD}(x, M)$ has zero area iff $x \in \partial(T(\mathcal{Y}_3))$.

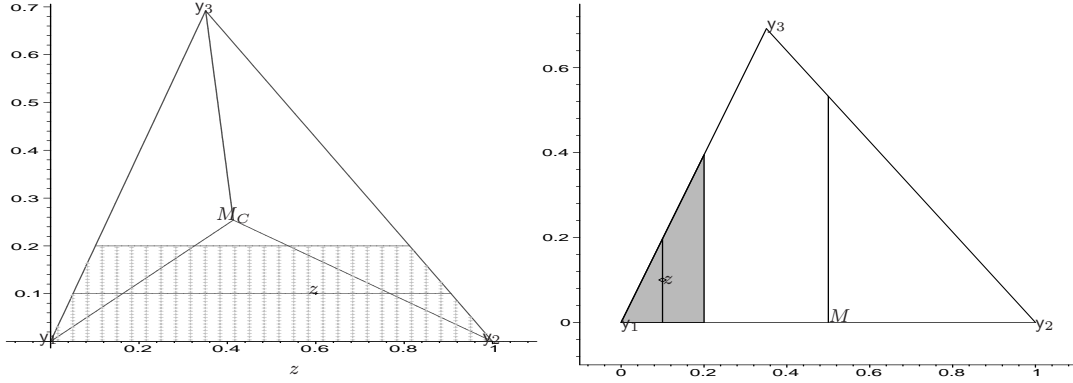


Figure 27: An example of directional-doubling proximity region (left) and double- X proximity region (right).

10.2 Double- X Proximity Maps

Without loss of generality, we can assume that $T(\mathcal{Y}_3) = T_b$. Partition the triangle T_b using the vertical line at $M_3 = (1/2, 0)$ as in Figure 27 (right). Let $R_1 := \{(x, y) \in T_b : x < 1/2\}$ and $R_2 := \{(x, y) \in T_b : x > 1/2\}$. If $(x, y) \in T_b$ with $x = 1/2$, assign (x, y) arbitrarily to one of R_1 or R_2 . We define the double- X proximity map as follows. For $z = (x_o, y_o) \in T_b \setminus \{y_1, y_2\}$

$$N_{DX}(z) := \begin{cases} \{(x, y) \in T_b : x \leq 2x_o\} & \text{if } z \in R_1, \\ \{(x, y) \in T_b : 1 - x \leq 2(1 - x_o)\} & \text{if } z \in R_2. \end{cases}$$

If $z = (x_o, y_o) \in \{y_1, y_2\}$, then $N_{DX}(z) := \{z\}$. See also Figure 27 (right). Notice that if $z \notin \{y_1, y_2\}$, then $N_{DD}(z, M)$ is a right triangle or a quadrilateral. Among the properties, **P1** and **P2** follows trivially. The vertical line at z divides the region into two pieces (half-way along the x -coordinate) so **P3** holds in this special sense. **P4** and **P5** fails to hold, since $N_{DD}(z, M)$ may be a quadrilateral for some $z \in T_b$. **P6** holds if $M = (1/2, y)$, otherwise $\mathcal{R}_S(N_{DX}, M)$ has positive area. **P7** also follows by definition. However, **P8** holds only when the regions R_1 and R_2 are constructed at a point where the vertical line divides the area into two equal pieces.

Property **P10** fails, since $N_{DX}(z)$ is constructed with the boundary of $T(\mathcal{Y}_3)$ and a line with a specific angle (perpendicular to the largest edge), by Corollary 5.6, geometry invariance for uniform data does not hold. That is, the distributions of relative arc density and the domination number of the corresponding PCD depend on the geometry of the triangle $T(\mathcal{Y}_3)$. Hence, it does not suffice to compute them for the standard equilateral triangle only, but instead one should compute them for each pair of (c_1, c_2) . Moreover, $\Lambda_0(N_{DX}) = \{y_1, y_2\}$ since $N_{DX}(x, M)$ has zero area iff $x \in \{y_1, y_2\}$.

11 Discussion and Conclusions

In this article, we discuss the construction of proximity catch digraphs (PCDs) in multiple dimensions. PCDs are a special type of proximity graphs which have applications in various fields. The class cover catch digraph (CCCD) is the first type of PCD family in literature (Priebe et al. (2001)) which is based on spherical proximity maps and has “nice properties” for uniform data in \mathbb{R} , in the sense that, the exact and asymptotic distribution of the domination number for CCCDs is available for one-dimensional uniform data. We determine some of the properties of the spherical proximity maps in \mathbb{R} (called *appealing properties*), and use them as guidelines for extending PCDs to higher dimensions. We also characterize the geometry invariance

for PCDs based on uniform data. Geometry invariance is important since it facilitates the computation of quantities (such as relative arc density or domination number) related to PCDs.

We discuss four PCD families in literature and introduce two new PCD families in this article. We investigate these PCD families in terms of the appealing properties and in particular geometry invariance for uniform data. We provide the asymptotic distribution of relative arc density and domination number for some of the PCD families. These tools have applications in spatial point pattern analysis and statistical pattern classification. We have demonstrated that the more the properties are satisfied, the better the asymptotic distribution of relative density. Furthermore, the availability of the asymptotic distribution of domination number also is highly correlated with the number of properties satisfied.

The spherical proximity regions were defined with (open) balls only, whereas the new proximity maps are not based on a particular geometric shape or a functional form; that is, the new proximity maps admit any type of region, e.g., circle (ball), arc slice, triangle, a convex or nonconvex polygon, etc. In this sense, the PCDs are defined in a more general setting compared to CCCD. On the other hand, the types of PCDs we introduce in this article are well-defined for points restricted to the convex hull of \mathcal{Y}_m , $\mathcal{C}_H(\mathcal{Y}_m)$. Moreover, the new families of proximity maps we introduce will yield closed regions. Furthermore, the CCCDs based on balls use proximity regions which are defined by the obvious metric, while the PCDs do not suggest an obvious metric.

The mechanism to define the proximity maps provided in this article can be also used for defining new (perhaps with better properties) proximity map families.

Acknowledgments

This research was supported by the research agency TUBITAK via the Kariyer Project # 107T647.

References

- Boots, B. N. (1986). Using angular properties of delaunay triangles to evaluate point patterns. *Geographical Analysis*, 18(3):250–260.
- Ceyhan, E. (2004). *An Investigation of Proximity Catch Digraphs in Delaunay Tessellations*. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218.
- Ceyhan, E. and Priebe, C. E. (2005). The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. *Statistics & Probability Letters*, 73:37–50.
- Ceyhan, E. and Priebe, C. E. (2007). On the distribution of the domination number of a new family of parametrized random digraphs. *Model Assisted Statistics and Applications*, 1(4):231–255.
- Ceyhan, E., Priebe, C. E., and Marchette, D. J. (2007). A new family of random graphs for testing spatial segregation. *Canadian Journal of Statistics*, 35(1):27–50.
- Ceyhan, E., Priebe, C. E., and Wierman, J. C. (2006). Relative density of the random r -factor proximity catch digraphs for testing spatial patterns of segregation and association. *Computational Statistics & Data Analysis*, 50(8):1925–1964.
- Chartrand, G. and Lesniak, L. (1996). *Graphs & Digraphs*. Chapman & Hall/CRC Press LLC, Florida.
- DeVinney, J. (2003). *The Class Cover Problem and its Applications in Pattern Recognition*. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218.

- DeVinney, J. and Priebe, C. E. (2006). A new family of proximity graphs: Class cover catch digraphs. *Discrete Applied Mathematics*, 154(14):1975–1982.
- DeVinney, J., Priebe, C. E., Marchette, D. J., and Socolinsky, D. (2002). Random walks and catch digraphs in classification. <http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf>. Proceedings of the 34th Symposium on the Interface: Computing Science and Statistics, Vol. 34.
- Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer Verlag, New York.
- Garfinkel, R. S. and Nemhauser, G. L. (1972). *Integer Programming*. John Wiley & Sons, New York.
- Janson, S., Łuczak, T., and Ruciński, A. (2000). *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York.
- Jaromczyk, J. W. and Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. *Proceedings of IEEE*, 80:1502–1517.
- Kimberling, C. (2008). Encyclopedia of triangle centers. <http://faculty.evansville.edu/ck6/encyclopedia/ETC.html>.
- Lee, C. (1998). Domination in digraphs. *Journal of Korean Mathematical Society*, 4:843–853.
- Lehmann, E. L. (1988). *Nonparametrics: Statistical Methods Based on Ranks*. Prentice-Hall, Upper Saddle River, NJ.
- Marchette, D. J. and Priebe, C. E. (2003). Characterizing the scale dimension of a high dimensional classification problem. *Pattern Recognition*, 36(1):45–60.
- Mardia, K. V., Edwards, R., and Puri, M. L. (1977). Analysis of central place theory. *Bulletin of International Statistical Institute*, 47:93–110.
- Okabe, A., Boots, B., and Sugihara, K. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley.
- Paterson, M. S. and Yao, F. F. (1992). On nearest neighbor graphs. In *Proceedings of 19th Int. Coll. Automata, Languages and Programming, Springer LNCS*, volume 623, pages 416–426.
- Priebe, C. E., DeVinney, J. G., and Marchette, D. J. (2001). On the distribution of the domination number of random class catch cover digraphs. *Statistics & Probability Letters*, 55:239–246.
- Priebe, C. E., Marchette, D. J., DeVinney, J., and Socolinsky, D. (2003a). Classification using class cover catch digraphs. *Journal of Classification*, 20(1):3–23.
- Priebe, C. E., Solka, J. L., Marchette, D. J., and Clark, B. T. (2003b). Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. *Computational Statistics & Data Analysis on Visualization*, 43-4:621–632.
- Schoenberg, F. P. (2002). Tessellations. *Encyclopedia of Environmetrics*, Wiley, NY, 3:2176–2179.
- Sen, M., Das, S., Roy, A., and West, D. (1989). Interval digraphs: An analogue of interval graphs. *Journal of Graph Theory*, 13:189–202.
- Toussaint, G. T. (1980). The relative neighborhood graph of a finite planar set. *Pattern Recognition*, 12(4):261–268.
- Tuza, Z. (1994). Inequalities for minimal covering sets in sets in set systems of given rank. *Discrete Applied Mathematics*, 51:187–195.

Weisstein, E. (2008). Triangle centers, Eric Weisstein's world of mathematics.
<http://mathworld.wolfram.com/TriangleCenter.html>.

West, D. B. (2001). *Introduction to Graph Theory*, 2nd Ed. Prentice Hall, N.J.